Modelling Dark Energy

Brendan Marc Jackson

Institute for Astronomy
School of Physics

THE UNIVERSITY
of EDINBURGH

University of Edinburgh
A dissertation presented in fulfilment of the requirements for the degree of Doctor of Philosophy at The University of Edinburgh

17 October 2011

Abstract

One of the most pressing, modern cosmological mysteries is the cause of the accelerated expansion of the universe. The energy density required to cause this large scale opposition to gravity is known to be both far in excess of the known matter content, and remarkably smooth and unclustered across the universe. While the most commonly accepted answer is that a cosmological constant is responsible, alternatives abound. This thesis is primarily concerned with such alternatives; both their theoretical nature and observational consequences.

In this thesis, we will dedicate Chapter 1 to a brief review on the fundamentals of general relativity, leading into the basics of theoretical cosmology. Following this we will recall some of the key observations that has lead to the standard ΛCDM cosmology.

The standard model has well known problems, many of which can be answered by the theoretical ideas of inflation. In Chapter 2 we explore these ideas, including a summary of classical field theory in the context of cosmology, upon which inflation is based. This also serves as the groundwork for Chapter 3, where the varied models of dark energy (and their motivations) are discussed - many of which are also reliant on field theory (such as quintessence).

These notions are combined in a model described in Chapter 4, where we describe our own addition to a scenario that unifies dark energy and inflation. This addition - involving a coupling of the inflation field to an additional one - alter the way reheating takes place after inflation, removing some of the shortcomings of the original proposal. The analysis is extended in Chapter 5, to include the effect of quantum corrections. There we show that although a cursory analysis indicates a coupling between quintessence and some other field does not necessarily give rise to dangerously large quantum corrections, provided the effects of decoupling are taken into account.

We move on in Chapter 6 to examine the basics of cosmological perturbation theory, and derive the general equations of motion for density and velocity perturbations for a system of fluids, allowing for the exchange of energy-momentum. We make use of this in Chapters 7 and 8, were we examine the growth of structure in a universe where

energy is exchanged between dark matter and dark energy. In particular, in Chapter 7 we see that a particular form of the interaction can lead to an instability in the early universe, and we derive the condition for this to be the case. In Chapter 8, we discuss how a similar interaction can lead to a mimicry of modified gravity, and relate this directly to cosmological observations.

Finally we summarise our conclusions and discuss avenues of future research in Chapter 9.

Declaration

This thesis is my own composition, and contains no material that has been accepted for the award of any other degree or professional qualification.

Parts of this thesis are based on researh published (or submitted for publication) seperately in refereed scientific journals. In each case, my role in this research was as primary contributor. More specific details follow.

Chapter 4 contains work presented in **Hybrid Quintessential Inflation**, by Mar Bastero-Gil, Arjun Berera, Brendan M. Jackson, Andy Taylor, published in Phys. Lett. B 678, 2:157-163, 2009. The analytic and numerical work were both primarily my contribution, while the initial ideas and introductory context was primarily the work of my collaborators.

Chapter 5 contains work presented in **Power suppression from disparate mass scales in effective scalar field theories of inflation and quintessence** by Mar Bastero-Gil, Arjun Berera, Brendan M. Jackson, published in Journal of Cosmology and Astroparticle Physics, 7, 010, 2011. I carried out the numerical and analytic work in the Results section, as well as making a significant contribution to the interpretation of the formalism in the rest of the document. The Appendices, as well as the introduction and description of the MDRS renormalisation scheme, were primarily the work of my collaborators.

Chapter 7 contains work presented in **On the large-scale instability in interacting dark energy and dark matter fluids** by Brendan M. Jackson, Andy Taylor and Arjun Berera published in Phys. Rev. D 79, 4:3526, 2009. The analytic and numerical calculations, as well as the interpretation of the results, were my own work, while my collaborators provided general project direction.

Chapter 8 contains work presented in **Unmodified Gravity** by Fergus Simp-

son, Brendan M. Jackson, John A. Peacock, published in MNRAS, 411, 2, 2011. The analytic calculations were my own, and I seperately confirmed the numerical calculations made, with the exception of the Fisher matrix analysis. I also made significant contributions to the interpretation of these results. The final discussion, as well as the introductory context, was the work of my collaborators.

Minor alterations from the published versions have been made, to provide greater exposition, as well as to maintain a consistent style.

Brendan Jackson
29 April, 2011

Acknowledgements

There are a number of people without whom this thesis would never have been completed.

First and foremost, I owe no small debt of gratitude for the support and patience of my supervisors: Andy Taylor and Arjun Berera. This research was only possible thanks to their guidance, and their encouragement was one of the key reasons I was able to see it through to the end.

I would also like to thank my friends and colleagues at the Royal Observatory. There are too many names to list them all, but I would particularly like to thank Alina Kiessling, Duncan Forgan, Hannah Durnall, Fergus Simpson, Julia Kennedy, Eric Tittley, Alan Heavens and John Peacock. Whether it was answering my questions about cosmology, mathematics or computers, cheering me up when I was feeling blue, or just letting me ramble on about my research, you all helped immensely.

To Paula Wilkie, John Barrow, Nathalie Dupin and Horst Meyerdierks I would also like to give a big thanks. The administrative and technical sides to my PhD were very smooth, and I know that keeping a department as busy and dynamic as the Institute for Astronomy running is no easy task.

Finally, I want so say thank you to my family. My parents have remained optmimistic and supportive throughout, and I hope I have made them proud. And last, but by no means least, a big thank you to my sister Bryony, for cooking me strange Swedish dishes and keeping me company during some of the more difficult times of writing this thesis.

‘This is the real secret of life – to be completely engaged with what you are doing in the here and now. And instead of calling it work, realise it is play.’
– **Alan Watts**

‘Physics is very muddled again at the moment; it is much too hard for me anyway, and I wish I were a movie comedian or something like that and had never heard anything about physics!’
– **Wolfgang Pauli**

Contents

Chapter 1

Friedmann-Lematre-Robertson-Walker Cosmology

1.1 General relativity

1.1.1 The force of gravity

We know of four fundamental forces of nature. Gravity was the first to be recognised, followed then by electromagnetism and eventually the strong and weak forces. The latter forces become rapidly weaker with increasing distance, leaving their discovery only to recent times. Electromagnetism is intrinsically stronger than gravity, but due to positive and negative charges being attractive to only their counterpart, large accumulations and positive or negative charge do not tend to occur in nature, even if the constituents of matter (such as the electrons and protons comprising the large abundances of hydrogen gas) are charged themselves.

On the other hand, all mass is attractive. An overdensity of matter simply attracts more mass to itself, increasing the overdensity and so increasing gravity's influence in a powerful feedback loop. Despite the weakness of the gravitational force, this effect is the dominant one. To understand the large scale behaviour of matter, one needs

to understand the large scale influence of the force of gravity; when constructing a theory of how matter behaves on the largest scales of all – cosmology – only a theory of gravity that is valid up to the largest scales will suffice. This theory is Einstein's general relativity.

We should recall some of the concepts and basic principles in general relativity. There are worse places to begin than by paraphrasing Wheeler's summary: matter tells spacetime how to curve, and the curvature of spacetime tells matter how to move. Let us unpack these statements. First, what does it mean for spacetime to be curved?

1.1.2 Flat and not-flat spacetime

Line elements

Consider two events in spacetime (instantaneous flashes of a lightbulb, if you like), separated by an infinitesimal *interval* $\mathrm{d}s$. What might this separation look like? In flat spacetime, we can use our knowledge of flat spatial geometry to tell us. We sum the squares of the distances. For flat spacetime rather than just space, we need to sum the square of the spatial parts, but subtract the time part (this is what distinguishes a four dimensional space from a three dimensional space with a single time dimension). Another term for this sum is the length of a line-element, and it looks as follows.

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\mathrm{d}t^2 \,. \tag{1.1}$$

The interval $\mathrm{d}s$ is an *invariant*; all observers will agree on this quantity, even though they might label their coordinate systems differently. The interval is related to a quantity with a more physical interpretation: the *proper time* τ, through

$$\mathrm{d}s^2 = -c^2\mathrm{d}\tau^2 \,. \tag{1.2}$$

If an observer were to move between two points in spacetime (defined by two sets of four coordinates each), the proper time τ separating these events would be the time experienced by that observer. The quantity τ for any observer will of course be agreed by everyone, as it too is an invariant. We should be at least a little pleased about this; how much time has passed for someone is a very physical thing. We can measure it with a clock or by counting grey hairs. We can all agree on this; we are reassured it is an aspect of reality we are measuring.

Relativity

Every observer has their own τ between points in spacetime. It is calculated from their own *paths* through spacetime between those points, and in general will not agree with the coordinate time t. Already this is different to the Newtonian picture, where if two observers measure two different passages of time, the inevitable conclusion would be that one was wrong. Here the conclusion is that the amount of time that has passed *really is different* for person to person (or spacetime path to spacetime path). The question 'how much time has passed?' becomes meaningless; the question 'how much time has passed for *this observer*?' is meaningful.

With an understanding of relativity we can calculate how much time should pass for each observer, given the geometry of spacetime (the metric) and a knowledge of their paths. Experimentally we can check by use of clocks, ranging from the biological (how many grey hairs has each observer acquired?) to the precise (how many atoms of an unstable element have decayed?)[1].

Another way of writing this line element would be to use spherical polars $(r, \theta, \phi$, along with time $t)$. In that form, again just using the familiar flat geometry:

$$\mathrm{d}s^2 = \mathrm{d}r^2 - c^2\mathrm{d}t^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\,. \tag{1.3}$$

Invariant quantities, such as $\mathrm{d}s$, provide a bedrock of objectivity in a theory were many other quantities are observer dependent. Much of the confusion in understanding the theory of relativity (special and general) comes from the fact that many quantities and statements are taken for granted as being entirely objective, when in fact they are not. They are only approximately true on small scales, at small velocities, in weak gravitational field. They are only approximately true in the sense of being cases of widespread subjective agreement. We have already seen this in the case of proper time. Asking 'how much time has passed?' makes sense on small scales, in weak gravitational fields, at small velocities, because everyone agrees. When these things are not true, it no longer makes sense to ask that question until an observer is specified ('how much time has passed *for you?*'). These problems will come back to haunt us in later discussions of cosmological perturbation theory.

One lesson that should be very quickly learned is that no great import should be

[1] Perhaps the sceptical reader, or one with a fondness for a Newtonian worldview of an objective clock, might assert that in such a disagreement, one observer really *is right*, and all the others wrong. In defense of relativity: if the passage of years does not mean the greying of hair, the decay of unstable elements or more generally, the ticking of clocks, what does it mean to say 'time has passed'?

given to the coordinates. Important quantities are not observer dependent; they are invariants and physically meaningful features of the world. Coordinates are not, and could indeed be labelled u, v, w, q for all the difference it would make to calculating observables. Usually though, the coordinates are *some observer's* coordinates. This can make them useful for quickly relating different observers' perspectives to the actual physical quantities.

Equation (1.1) is only valid in flat spacetime, just like Pythagoras' theorem is only valid on flat geometries. This gives a little insight as to what curved spacetime means – it is not flat, and flat spacetimes obey equation (1.1). Still, an insight about what something is *not* is not really much insight at all, so we should look at an example of a curved spacetime.

Curved spacetime

This is the Schwarzschild metric.

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{rc^2}\right)^{-1} \mathrm{d}r^2 - \left(1 - \frac{2GM}{rc^2}\right) c^2 \mathrm{d}t^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2)\,. \tag{1.4}$$

The coordinate system is spherical polars, along with a time coordinate. It happens to be the spacetime produced by a point-like object of mass M situated at $r = 0$ (for all t). A few interesting points of note emerge from looking at equation (1.4).

First, the spherical symmetry is apparent as the angular part of the line element is unchanged from how it would be in flat space-time, as is quickly verified by comparison with equation (1.3).

Second, the interval between two events separated only by $\mathrm{d}r$ or $\mathrm{d}t$ is no longer the flat spacetime result. Instead there is a prefactor that depends upon the radial coordinate. The effect of the curvature of spacetime has changed the interval (and so the proper time). Two observers at different, but fixed, radial values will experience a different amount of time passing in a given amount of coordinate time. Does this matter? Yes, it does. Coordinate time is the proper time of *someone*: at infinite radial distance, the coordinate time approximates the proper time, as the prefactors tend to unity. The coordinate time is the proper time of an observer an infinite distance away.

Third, when the radial coordinate takes the value of $r = r_s \equiv 2GM/c^2$, the Schwarzschild radius, there is a singularity. This is only an apparent singularity due to a poor choice of coordinates, and can be removed with a more sensible choice. It does signal an interesting place in the metric; light emitted from inside this event horizon can never be seen by an observer that remains outside the event horizon. This is

not readily apparent, however, because we have yet to discuss the rest of Wheeler's summary: how does spacetime tell matter how to move? For this, we should look to the Principle of Equivalence, and see how this naturally introduces the mathematics of tensors.

1.1.3 The Principle of Equivalence

In Newtonian physics, the mass of an object plays two distinct roles. It serves to provide a measure of inertia; a quantitative description of how much force needs to be applied to an object to accelerate it by a set amount. Along with this inertial mass, there is also a gravitational mass – how strong a gravitational force is exerted and felt by the object. It is an unexplained feature of the theory that the inertial and gravitational masses of an object are numerically equal.

Because of this equivalence of inertial and gravitational mass, Einstein famously realised that an observer could not distinguish in any way, through any conceivable experiment, between free-fall through a homogeneous gravitational field, and the effect of acceleration. The equations of Newton look identical in each case. Of course, gravitational fields are generally not homogeneous, since the gravitational force drops as the inverse square of the distance. Still, over a sufficiently small region of spacetime it can always be regarded as constant. Taking this notion as fundamental, the Principle of Equivalence can be formulated thus [148]: *at every point in an arbitrary gravitational field it is possible to choose a 'locally inertial coordinate system' such that, within a sufficiently small region of the point in question, the laws of nature take the same form as in unaccelerated Cartesian coordinate systems in the absence of gravitation.*

This statement remarkably implies a great deal of the contents of general relativity. For a particle freely falling under purely gravitational forces, there is a coordinate system where the particle moves in straight lines,

$$\frac{\mathrm{d}^2\zeta^\alpha}{\mathrm{d}\tau^2} = 0\,, \tag{1.5}$$

where τ is the proper time of the particle. Here the superscript α is a *label* (not an exponent). This labels components of four vector ζ^α. Equation (1.5) is actually four equations (with α ranging from 0 to 3).

Since it is a Cartesian coordinate system, the proper time should match equation (1.1). We can write it more succinctly with

$$c^2\mathrm{d}\tau^2 = \eta_{\mu\nu}\mathrm{d}\zeta^\mu\mathrm{d}\zeta^\nu\,. \tag{1.6}$$

The new object $\eta_{\mu\nu}$ is the diagonal matrix diag$(1, -1, -1, -1)$, with μ and ν serving to label each of the sixteen components. This is also the first encounter with the *Einstein summation convention* – repeated indices in the above equation (in this case μ and ν) are summed over and range from 0 to d, where d is the number of spatial dimensions. For four dimensional spacetime (one time, three spatial dimensions), this is from 0 to 3. The four ζ^μ are each of the coordinates. In a Cartesian coordinate system we use $\zeta^0 = ct$, $\zeta^1 = x$, $\zeta^2 = y$, and so on. Expanding out equation (1.6 just gives the explicit Minkowski metric of equation (1.1).

General coordinate systems

Now consider a general coordinate system x^μ. This may be at rest with respect to the particle, or accelerated, rotating – anything at all. We treat the freely falling coordinates ζ^μ as functions of the x^μ. In this way, we are relating a general coordinate system x^μ across a potentially large region of spacetime (where the behaviour of particles is as yet unknown) to a description of spacetime at each point as locally Minkowski, with particles obeying equation (1.5). Expanding out the functions $\zeta^\mu(x^\mu)$ using partial derivatives, it is clear that

$$0 = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial\zeta^\alpha}{\partial x^\mu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\right) = \frac{\partial\zeta^\alpha}{\partial x^\mu}\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\tau^2} + \frac{\partial^2\zeta^\alpha}{\partial x^\mu\partial x^\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} \tag{1.7}$$

Following Weinberg, we can multiply this by $\partial x^\lambda/\partial\zeta^\alpha$, and recall the product rule,

$$\frac{\partial\zeta^\alpha}{\partial x^\mu}\frac{\partial x^\lambda}{\partial\zeta^\alpha} = \delta^\lambda_\mu\,, \tag{1.8}$$

where $\delta^\lambda{}_\mu$ is the Kronecker-delta tensor (zero unless both indices match, when instead it is equal to unity). The resulting equation is the equation of motion for particles in any coordinate system in general relativity,

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0\,. \tag{1.9}$$

The object $\Gamma^\mu_{\alpha\beta}$ is the *affine connection* and defined only in terms of coefficients that transform from locally inertial coordinates to general coordinates,

$$\Gamma^\mu_{\alpha\beta} = \frac{\partial x^\mu}{\partial\zeta^\nu}\frac{\partial^2\zeta^\nu}{\partial x^\alpha\partial x^\beta}\,. \tag{1.10}$$

Note the symmetry under exchange of the lower two indices; the order of the partial differentiation does not matter.

The metric tensor

Performing similar surgery to the line-element allows this too to be described in arbitrary coordinates. Once more, local to a point, the line-element will look like that of flat spacetime in appropriately chosen free fall coordinates (ζ^μ). In arbitrary coordinates, it takes the form of

$$\mathrm{d}s^2 = -c^2\mathrm{d}\tau^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu\,. \tag{1.11}$$

The new object $g_{\mu\nu}$ is the *metric tensor*, describes how the line-element behaves as we move around this arbitrary coordinate system. Like the affine connection, it is so far defined in terms of the transformation coefficients,

$$g_{\mu\nu} = \frac{\partial\zeta^\alpha}{\partial x^\mu}\frac{\partial\zeta^\beta}{\partial x^\nu}\eta_{\alpha\beta}\,. \tag{1.12}$$

The metric tensor is an extremely important object. Given an arbitrary coordinate system, it (along with the affine connection) defines the locally inertial coordinates local to any point, that must exist to satisfy the Principle of Equivalence. The affine connection can be related to the metric tensor, by taking partial derivatives of $g_{\mu\nu}$. Defining the inverse of $g_{\mu\nu}$ to be $g^{\mu\nu}$ (in the sense that $g_{\alpha\nu}g^{\alpha\mu} = \delta^\mu{}_\nu$, then

$$\Gamma^\alpha_{\lambda\mu} = \frac{1}{2}g^{\alpha\nu}\left(\frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\nu}\right)\,. \tag{1.13}$$

The metric tensor could therefore be regarded as fundamental, with the affine connections related to gradients of its components.

The Newtonian limit

How should we interpret the metric tensor? From looking at how it appears in our equation for the line element, we see it defines the geometry of spacetime; deviations from the Minkowski metric of $\mathrm{diag}(1,-1,-1,-1)$ signals curved spacetime. We should therefore interpret deviations in the metric tensor from flat spacetime to being somehow related to the gravitational field. We can gain a little more physical insight by making contact with the Newtonian limit.

Consider a slowly moving particle in a weak, stationary gravitational field. We can neglect $\mathrm{d}\mathbf{x}/\mathrm{d}\tau$ with respect to $\mathrm{d}t/\mathrm{d}\tau$ and approximate (1.9) as,

$$\frac{\mathrm{d}^2x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{00}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 = 0\,. \tag{1.14}$$

Furthermore, since the field is stationary, we can expect the metric tensor to also have no time derivatives. This leaves

$$\Gamma^{\mu}_{00} = -\frac{1}{2} g^{\mu\nu} \frac{\partial g_{00}}{\partial x^{\nu}} \,. \tag{1.15}$$

We can also make an expansion about Minkowski spacetime, since we claim the field is only weak. So we use

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \,, \tag{1.16}$$

where we work only to first order in $h_{\mu\nu}$.

With the affine connection computed, the equations of motion can be written in terms of this perturbed Minkowski metric,

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}\tau^2} = \frac{1}{2} c^2 \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \nabla h_{00} \,. \tag{1.17}$$

$$\frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} = 0 \,. \tag{1.18}$$

The second equation shows $\mathrm{d}t/\mathrm{d}\tau$ is constant. Dividing the first equation through by this, we find

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = \frac{1}{2} c^2 \nabla h_{00} \,. \tag{1.19}$$

Comparing to the Newtonian result for the acceleration due to a massive body,

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = -\nabla\phi \,, \tag{1.20}$$

with $\phi = GM/r$ then it becomes clear $h_{00} = -GM/c^2 r$. We conclude the components of the metric tensor act as potentials.

1.1.4 Tensors

The proper time, or the interval, are invariants. They are the same for all observers, no matter the coordinate system. If we have $\mathrm{d}s^2 = g_{\alpha\beta}\mathrm{d}x^{\alpha}\mathrm{d}x^{\beta}$ and change to a new coordinate system x', we can write $\mathrm{d}x^{\mu}$ in terms of $\mathrm{d}x'^{\mu}$. Substituting and *demanding* the resulting equation has the form $\mathrm{d}s^2 = g'_{\alpha\beta}\mathrm{d}x'^{\alpha}\mathrm{d}x'^{\beta}$, we can deduce the transformation law for a tensor:

$$g'_{\alpha\beta} = \frac{\partial x^{\mu}}{\partial x'^{\alpha}} \frac{\partial x^{\nu}}{x'^{\beta}} g_{\mu\nu} \,. \tag{1.21}$$

The definition of $g_{\mu\nu}$ in terms of $\eta_{\mu\nu}$ was an example of this. The components of $g_{\mu\nu}$ change in just such a way as to preserve the form of the equation when the coordinate system is changed, and the four-vectors $\mathrm{d}x^{\mu}$ also change.

Contravariant and covariant tensors

We can define a new vector

$$\mathrm{d}x_\nu \equiv g_{\alpha\nu}\mathrm{d}x^\alpha\,, \tag{1.22}$$

which gives the metric simply as

$$\mathrm{d}s^2 = \mathrm{d}x_\nu \mathrm{d}x^\nu\,. \tag{1.23}$$

This defines the relationship between the *contravariant* components of a 4-vector A^μ and the *covariant* components A_μ. The names reflect how the components transform with respect to a change in coordinates. The transformation laws are

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu\,, \tag{1.24}$$

$$A'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} A_\nu\,. \tag{1.25}$$

These *define* the notion of a tensor; they are objects that transform appropriately under a change of coordinates. The definition can be generalised to any tensor, by multiplying by an appropriate factor for each index. This also applies to tensors of mixed components. These are written with some indices at the top, and some at the bottom, like $T^\mu{}_\nu$.

Not all objects with multiple indices are tensors. The affine connections ($\Gamma^{\alpha\beta}_\gamma$), despite appearances, are not tensors, as they do not transform appropriately. Another object that is not normally a tensor is the partial derivative of a tensor, like $\partial_\mu A^\nu \equiv \partial A^\nu/\partial x^\mu$.

One sort of tensor has a very simple transformation law; these are tensors without any indices at all, such as the interval $\mathrm{d}s^2$. These are called *scalars*, and have the delightful property of being identical in all coordinate systems (and so for all observers). These invariants form the crucial observational bedrock. One way of constructing an invariant is by *contracting* two four vectors, e.g. $A^\mu A_\mu$ to produce a scalar.

General covariance

Equations built entirely out of tensors are true in any coordinate system, purely from the definition of a tensor[2]. Such equations are said to be *generally covariant*, which has nothing to do with the distinction between contravariant and covariant components and is just one of many examples of poor nomenclature in this subject.

[2]The usefulness of this feature is readily apparent after recalling the tedium that occurs when writing down the correct form of Newton's Laws in a non-inertial coordinate system

The fact that the partial derivative of a tensor is not itself a tensor is particularly troubling. It would be nice to know how to take a derivative of a tensor in a covariant fashion. The way forward is to define a covariant derivative ∇_μ which is the partial derivative, *plus* new terms that 'correct' for the wrong transformation properties under coordinate changes. The general rule for a four-vector is

$$\nabla_\nu A^\mu \equiv \partial_\nu A^\mu + \Gamma^\mu_{\alpha\nu} V^\alpha \,. \tag{1.26}$$

Similar rules exist for the tensors with more indices.

From special to general relativity

In special relativity, the derivatives of the metric tensor vanish, and so all the affine connections are zero. As might be expected, covariant derivatives are just partial derivatives. Now we make the following argument from general covariance - if a tensor equation is true in one coordinate system, it is true in all of them.

Special relativity is general relativity in a particular (Minkowski) spacetime, therefore provided we can write the equations of special relativity in tensorial form, we have a bootstrap method for writing down the equations of general relativity by referring to special relativity. As most equations in special relativity come pre-assembled in the form of four-vectors, this is often as simple as replacing the partial derivatives with covariant ones – these remain true statements in special relativity, and therefore are also true in general relativity.

Unfortunately, this method is not without some ambiguity. Additional tensors do exist that vanish in special relativity, and so can be introduced into nearly any equation of special relativity without damaging the Minkowskian result. Experiment is the ultimate arbiter in such cases; simplicity is often used as a working principle for the sake of progress.

1.1.5 Spacetime tells matter how to move

In the Newtonian world we are familiar with, an object without a force acting upon it moves in a straight line. Gravity acts as a force that stops things behaving in this way, sending them in apparently strange paths (like the orbits of planets around a star).

In general relativity, gravity is no longer written as a force. Instead it is a change to the geometry of spacetime, through which objects move. Flat spacetime has no gravity; objects should still move in straight lines assuming no other forces are present.

We should recall what a straight line *is*, with a suitably general definition that we can use it when the geometry is no longer flat.

A straight line is the shortest path between two points. This should be treated as a definition, not (as it is often presented) a conclusion. To be technically precise, we should also note that the longest path between two points could also be a straight line (as simple examples, think of flat space where coordinates wrap around, such as the surface of a torus or the early computer game *Asteroids*). We should generalise then to a straight line being the path that extremises the distance between two points. These sorts of paths are known as *geodesics*.

This definition is useful, as it holds even in non-flat space. In a curved spacetime, we will say that a straight line is the path that extremises the interval (or proper time) between two points. The curvature of spacetime tells matter how to move by defining what these geodesics are. With no forces acting upon a particle, it will follow a geodesic. Gravity emerges not as a force at all, but as a change to the background spacetime geometry. In this respect a freely falling skydiver, or a planet orbiting a star, should not be regarded as having a force acting upon it. Rather these are examples of objects that are freely moving along the straightest lines they can follow, given the curvature of spacetime.

1.1.6 Matter curves spacetime

Now we understand what it means for spacetime to be curved, and the implications this has on the movement of matter, we can turn to the final part of Wheeler's summary of general relativity. How does matter curve spacetime?

Ideally we need a way of relating the curvature of spacetime, characterised by $g_{\mu\nu}$ to the distribution of matter. In Newtonian cosmology only the distribution of mass is important, but in relativity the notion of a mass becomes clouded by the notion of a 'relativistic mass'. There is no way to properly derive the field equations regardless, so we will make the leap of reasoning and simply guess that all forms of energy will gravitate (justified by the experimental success).

The energy-momentum tensor

Energy and momentum are conserved. We would therefore like four conservation laws. A single four-vector sadly only gives one conservation law. As an example, the four-current in electromagnetism J^μ is conserved, describing conservation of charge through

$\nabla_\mu J^\mu$. Just one conservation law is not enough for energy and the three components of momentum, so we write down something sensible that does have enough components:

$$\nabla_\mu T^{\mu\nu} = 0\,, \tag{1.27}$$

where $T^{\mu\nu}$ is a tensor referred to as the energy-momentum (or sometimes stress-energy) tensor. The components have a physical interpretation: T^{00} is $c^2\times$ mass density = energy density; T^{12} is the x-component of current of y-momentum, and so on. The physical meaning also demands that $T^{\mu\nu}$ is another symmetric tensor.

For a perfect fluid, the energy-momentum tensor has a particular form, using ρ for energy density and P for pressure,

$$T^{\mu\nu} = \left(\rho + P/c^2\right) U^\mu U^\nu - P g^{\mu\nu}\,, \tag{1.28}$$

where U^μ is the four-velocity of the fluid. This reduces to the special relativity result of $\mathrm{diag}(c^2\rho, -P, -P, -P)$ in the appropriate limit.

In Newtonian gravity, the second derivative of some potential ϕ is related to the matter distribution directly (Poisson's equation). We have already seen that in the weak field limit, this potential is proportional to g_{00} so we should expect our new potential $G^{\mu\nu}$ to be somehow related to second derivatives of the metric tensor. We might guess it takes the following form:

$$G_{\mu\nu} = -\frac{8\pi G}{c^4} T_{\mu\nu}\,. \tag{1.29}$$

Constructing the field equations

Remarkably, there is in fact a unique tensor that is *linear* in second derivatives of the metric (see p. 133 of Weinberg [148] for the proof), and contains nothing higher than those second derivatives. This is the *Riemann tensor*:

$$R^\mu{}_{\alpha\beta\gamma} = \frac{\partial \Gamma^\mu_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial \Gamma^\mu_{\alpha\beta}}{\partial x^\gamma} + \Gamma^\mu_{\sigma\beta}\Gamma^\sigma_{\gamma\alpha} - \Gamma^\mu_{\sigma\gamma}\Gamma^\sigma_{\beta\alpha}\,. \tag{1.30}$$

The Riemann tensor can be contracted to the *Ricci tensor* and further to the *Ricci scalar*,

$$R_{\alpha\beta} = R^\mu{}_{\alpha\beta\mu}\,, \quad R = R_\mu{}^\mu = g^{\mu\nu} R_{\mu\nu}\,. \tag{1.31}$$

With these new tensors, we are in a position to work out what $G_{\mu\nu}$ is. There are five criteria that must be satisfied:

1. $G_{\mu\nu}$ is a tensor.

2. We assume $G_{\mu\nu}$ contains terms with two derivatives of the metric tensor (linear in second derivatives, or quadratic in first derivatives).

3. $T_{\mu\nu}$ is conserved (vanishes under a covariant derivative), therefore $G_{\mu\nu}$ must also be.

4. $T_{\mu\nu}$ is symmetric, so $G_{\mu\nu}$ must be.

5. For a weak stationary field, we return to the Newtonian limit $G_{00} \sim \nabla^2 g_{00}$.

From this, it is possible to work out what $G_{\mu\nu}$ must be. We shall leave the arguments in Weinberg, and simply quote the result:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \,, \tag{1.32}$$

yielding the *Einstein Field Equations*,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -8\pi G T_{\mu\nu} \,. \tag{1.33}$$

This is the key equation to relate curvature to the distribution of energy-momentum.

One final note, that will become important later. If we relax the second assumption and allow $G_{\mu\nu}$ to contain terms with fewer than two derivatives of the metric, an extra term is possible. This is just a constant multiplied by $g_{\mu\nu}$.

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \lambda g_{\mu\nu} = -8\pi G T_{\mu\nu} \,. \tag{1.34}$$

This term was originally introduced by Einstein in the spurious hope it would provide a steady-state cosmology (the resulting solution is unstable, and so of no use). History still lingers in the name of the constant λ, referred to now as the *cosmological constant*. The term has found itself reappearing in our models of cosmology, and will be discussed in detail later on. For now, the important point to take away is that this term could equally be viewed as a contribution to the *other* side of the equation, by writing $\lambda g_{\mu\nu}$ in the form of a an energy momentum-tensor.

What sort of energy-momentum tensor would it have? Look at equation (1.28) again. Clearly a perfect fluid has precisely the form of $T_{\mu\nu} \propto g_{\mu\nu}$ if $\rho c^2 = -p$ *and* the energy density (and so the pressure) is constant. To make this point more emphatically before moving on: *the simplest modification to the field equations, a cosmological constant, is indistinguishable from a smooth perfect fluid with constant* $\rho c^2 = -p$ *permeating throughout all of spacetime.*

1.2 Cosmology - theory

We would like to solve Einstein's field equations for the entire universe, all at once. This is ambitious, but with some simplifying assumptions, possible.

Principles

Usually one of the first steps in simplifying a physical problem to the level of solubility is to identify any symmetries in the problem. This is precisely the approach to take here. By making the assumption the universe is isotropic – that is appears the same no matter in which direction we look our from the Earth – we can immediately infer spherical symmetry will be a key ingredient. We can also guess that using spherical polar coordinates will be a sensible choice to make.

Ideally, our location in the universe should not be special. This notion is elevated to a principle, named the *Copernican principle* in a reference to our history in the understanding of the solar system. In the case of our discussed cosmology, this suggests any observer would view the universe as isotropic about their location. By assuming the Copernican principle and isotropy about our present location, it necessarily follows the universe is homogeneous, as well as isotropic about every observer. Often the hypothesis of a spatially homogeneous and isotropic universe is labelled as the Cosmological Principle, and taken as a starting premise by itself. Either way, such assumptions severely restrict the form cosmological models might take.

It should be noted that these assumptions are clearly wrong on small scales. The universe is obviously not homogeneous or isotropic out as far as our neighbouring galaxies, or indeed past scales of clusters of galaxies. It is hoped, and to some extent has been confirmed, that on the largest scales the universe does average out to give this homogeneous, isotropic universe.

Cosmic time

To define a time coordinate in such a universe, it is quite plausible to use the universe itself as a clock. Since the universe is homogeneous, global properties exist such as the density of matter. Observers can agree to start their clocks when the global density reaches a certain value (assuming such a function is everywhere decreasing or increasing monotonically), through an exchange of light signals. The end result is a *cosmological time* coordinate. All observers agree upon this, provided the history of the universe appears identical also, in the sense that two observers with coordinates x'^{μ} and x^{μ} will

find that at any coordinate point y^μ

$$g_{\mu\nu}(y) = g'_{\mu\nu}(y), \tag{1.35}$$

$$T_{\mu\nu}(y) = T'_{\mu\nu}(y), \tag{1.36}$$

and so on for all the cosmological tensors. Observers that satisfy this property are called *fundamental observers*. Cosmologists living in typical galaxies might be hoped to be a practical example of fundamental observers, provided they are not moving dramatically with respect to the local matter density (generally that are not). It is also now clear why they all agree on a cosmological time, for recalling the cosmological density ρ is a scalar then $\rho(t') = \rho(t)$. Thus $t = t'$, and so all fundamental observers use the same, cosmological time.

Robertson-Walker spacetime

With this, we can write down the following form for the metric of such a spacetime,

$$c^2\mathrm{d}\tau^2 = c^2\mathrm{d}t^2 - R^2(t)\left[f^2(r)\mathrm{d}r^2 + g^2(r)\mathrm{d}\psi^2\right] . \tag{1.37}$$

where $\mathrm{d}\psi^2$ is the angular part of the metric ($\mathrm{d}\psi^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$). The functions $f^2(r)$ and $g^2(r)$ are unconstrained at this point, but there is freedom to choose $g = r^2$ to make it look as close to a flat spacetime as possible.

To progress further requires some study of maximally symmetric spaces (again see [148]), so we skip to the end result: the most general form of metric that satisfies our cosmological assumptions,

$$c^2\mathrm{d}\tau^2 = c^2\mathrm{d}t^2 - R^2(t)\left\{\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\right\} . \tag{1.38}$$

This is known as the *Robertson-Walker metric*. The cosmology produced by this metric is Friedmann-Lematre-Robertson Walker (FLRW) cosmology.

Geometry

The function $R(t)$ remains an unknown function of time, while k is a constant that with a suitable choice of units can be chosen to have the value $+1$, 0 or -1. Each of these choices is a different geometry for the spatial part of the metric. The choice of a flat space (not spacetime) is $k = 0$. The values $k = 1$ and $k = -1$ are choices of positive and negative curvature respectively, sometimes referred to as closed and open universes. These are the other two spatial geometries that can produce a maximally symmetric spacetime within the assumptions we have made.

Expansion and proper distance

One of the curious features of the Robertson-Walker metric is that it appears fundamental observers, who are *at rest with respect to the coordinate system*, will still become further separated or closer together depending upon how the function $R(t)$ grows or declines. As in practice $R(t)$ is growing, this effect is known as the *expansion* of the universe.

A useful way of characterising this is to consider the *proper distance* between two observers. This is the distance between two observers at separate r coordinates *at the same cosmic time* t. Obviously we need to use the metric to work out the proper distance $d_{\text{prop}} = \sqrt{\mathrm{d}s^2} = \sqrt{-\mathrm{d}c^2\tau^2}$. For small enough separations, this gives

$$d_{\text{prop}}(t) = R(t) \int \frac{\mathrm{d}r}{\sqrt{1-kr^2}} \approx R(t) r \,, \tag{1.39}$$

although at larger distances the curvature of the universe can no longer be neglected, and this relation can break down. Then we would need to do the integral properly, as well as grapple with the physical interpretation of a radial coordinate whose meaning rapidly deteriorates from its Newtonian one.

Historically, the effect of the expansion was first discovered through observations of galaxies made by Slipher, that showed anomalously large number of galaxies with their absorption features shifted toward the redder part of the spectrum. A few years later, Hubble made this point more forcefully. This *red shift* was attributed to the Doppler effect from galaxies receding from us, although we this is neither technically true, nor a helpful interpretation.

To see how the Robertson-Walker metric predicts this redshift, consider the path of light-rays. Light always moves at c, no matter the coordinate system. This means light rays must follow geodesics where $\mathrm{d}\tau = 0$ (called null geodesics), or

$$0 = c^2\mathrm{d}\tau^2 = c^2\mathrm{d}t^2 - R^2(t)\frac{\mathrm{d}r^2}{1-kr^2} \,. \tag{1.40}$$

Thus, we see that

$$\int \frac{\mathrm{d}t}{R(t)} = f(r) \equiv \int \frac{\mathrm{d}r}{\sqrt{1-kr^2}} \,. \tag{1.41}$$

For a photon emitted from one galaxy to another, the domains of integration range from t_0 when the photon is emitted, to t_1 when the photon is received (and likewise, from r_0 to r_1 where these take place). The function $f(r)$ is constant, as galaxies (as fundamental observers) have constant comoving coordinates. A photon emitted at a later time will be received at a later time, but as $f(r)$ is constant then the integral

cannot be altered. This forces the condition $\mathrm{d}t_{\mathrm{emit}}/\mathrm{d}t_{\mathrm{obs}} = R(t_{\mathrm{emit}})/R(t_{\mathrm{obs}})$. This reasoning applies as much as to the wavecrest of light as to the emission of separate photons, and we have a relation between the ratios of emitted and observed frequency of light ν and the scale-factor:

$$\frac{\nu_{\mathrm{emit}}}{\nu_{\mathrm{obs}}} \equiv 1 + z = \frac{R(t_{\mathrm{obs}})}{R(t_{\mathrm{emit}})}\,. \tag{1.42}$$

Typically z is the quantity referred to as redshift. It is produced by the fact distant galaxies are time dilated by the global cosmological gravitational field. It should also be clear that this has nothing to do with any form of Doppler shifting.

If we consider how the proper distance changes with time for a nearby galaxy, using equation (1.39) (and with only a little abuse of language, call it a velocity v) then,

$$v = \frac{\dot{R}}{R}\, d_{\mathrm{prop}}\,, \tag{1.43}$$

at least for small distances. Hubble interpreted the redshift of distant galaxies as due to a velocity, and noted this velocity was proportional to the distance. The above is just such the relation, with Hubble's 'constant' $H = \dot{R}/R$ - though this is obviously not a constant, after all.

Scale-factor equations

It would be nice to have an equation that described the evolution of R. So far, the Einstein field equations have not yet been used. The Robertson-Walker metric should of course satisfy these, using an appropriate energy-momentum tensor. Using the energy-momentum density of a perfect fluid is sufficient, with $T^{\mu\nu} = diag(c^2\rho, -p, -p, -p)$. The density ρ and pressure p are of course homogeneous (functions only of time, not space).

The proof that this metric satisfies the field equations for a homogeneous, uniform distribution of matter is to plug the metric into the field equations directly. There are sufficient explicit derivations later in this document to make skipping this one an appropriate decision. The result of such an exercise, however, yields two equations for the evolution of the scale-factor R. These are (with dots indicating time derivatives),

$$H^2 = \frac{\dot{R}^2}{R^2} = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2}\,, \tag{1.44}$$

known as *Friedmann's equation* and

$$c^2\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}\left(\rho c^2 + 3P\right)\,, \tag{1.45}$$

which is the *acceleration equation*.

For a flat universe, $k = 0$. Then from the form of these two equations, the dimensions of $R(t)$ become irrelevant. Usually $R(t)$ is scaled by the size of the scale-factor today, R_0 (a subscript 0 will usually indicate a quantity today, such as today's time t_0). The dimensionless scale-factor is denoted $a(t)$, with $a_0 = 1$ if we scale it appropriately.

If the universe is indeed flat, the fact $k = 0$ implies a relationship between the density and the expansion rate. At every epoch, the total density matches the *critical density* ρ_c:

$$\rho_c = \frac{3H^2}{8\pi G}\,. \tag{1.46}$$

A universe that is closed ($k = 1$) has a density above this value; a universe that is open ($k = -1$) has the opposite. This inspires us to define a *density parameter* as the ratio between the actual density and this critical density,

$$\Omega(t) \equiv \frac{\rho}{\rho_c} = \frac{8\pi G\rho}{3H^2}\,. \tag{1.47}$$

Knowing the value of $\Omega(t)$ (at any time) imparts the knowledge of the spatial geometry of the universe. If it is unity (for a flat universe), it remains unity for all time. If notation were consistent throughout the subject, Ω_0 would be used to indicate today's density parameter. Instead, to save ourselves from an overabundance of subscripts, it is often simply written as Ω, with the time dependence explicitly noted when that version of the symbol is used.

The importance of being flat

A great deal of importance has been placed upon the flatness of the universe in the past, mainly because it was believed to be the major deciding factor in the fate of the universe. A simple physical interpretation of the geometry term in Friedmann's equation is that of the 'total energy' of the universe; the derivative term can be viewed as the 'kinetic' energy, while the density term is the potential. Viewed this way, the behaviour of the universe can be thought of as an initial 'kick' imparting an appropriate expansion, that is gradually slowed by a (presumably ever decreasing) gravitational tug. This simple kinematic picture is nearly correct. Assuming no unusual matter content of the universe a closed universe is bound and does indeed recollapse; an open universe is unbound and expands forever. The flat case is the border between these two cases, barely expanding forever.

Unfortunately, the universe does appear to have some unusual matter content (or at least, for some reason, the gravitational effect does not appear to be getting weaker

with time in the way we expect). This decouples the fate of the universe from its geometry. Furthermore, observations appear roughly consistent with a flat universe, and there is some theoretical prejudice that this should be the case. For this reason, most of the time we will be only considering a universe that is flat (or sufficiently close to being so that we can neglect this aspect).

1.2.1 Evolution of matter

Continuity equation

Before solving Friedmann's equation, it helps to know how the density ρ evolves with scale-factor. For that, we make use of the fact that the energy-momentum tensor is conserved: $\nabla_\mu T^{\mu\nu} = 0$. This leads to a continuity equation,

$$\dot{\rho} + 3\frac{\dot{R}}{R}\left(\rho + \frac{P}{c^2}\right) = 0\,. \tag{1.48}$$

Often the relationship between pressure and density is assumed to be that of proportionality, with the constant of proportionality w defined as $w \equiv p/(c^2\rho)$. The above equation is simple enough to integrate directly if w is assumed to be constant,

$$\rho \propto R^{-3(1+w)}\,. \tag{1.49}$$

A cosmic inventory

We should now list the things we expect to find in the universe, and discuss how they evolve in an expanding universe.

- *Matter* – sometimes called 'dust', this is pressureless matter. Intuitively, a collection of particles in a set volume would have their density decrease proportional as the volume increases. Treating the volume as the cube of the scale-factor, the result for pressureless ($w = 0$) matter agrees with intuition, $\rho_{\rm m} \propto R^{-3}$.

- *Radiation* – relativistic matter is not pressureless; recall that light exerts a pressure proportional to its energy density, with constant of proportionality $w = 1/3$. The result is $\rho_{\rm r} \propto R^{-4}$ and can be neatly understood when considering how an electromagnetic wave is stretched by the expansion. The stretched wavelength of course reduces the energy, proportionally to the stretching, in addition to the volume dilution of the waves.

- *Neutrinos* – for relativistic particle, rather than waves, the result can also be understood as a reduction in the momentum with the expansion, proportional to the scale-factor. For non-relativistic matter, this is a negligible energy loss (most of the energy per particle is in the rest mass), while for relativistic matter this is not so. For neutrinos (or other relativistic particles), the behaviour is like radiation, $w = 1/3$ and $\rho_\nu \propto R^{-4}$.

There is also the possibility of a cosmological constant, either as a vacuum energy or a modification to gravity. We should add this to our list.

- *Cosmological constant* – recall this is indistinguishable from a smooth fluid permeating the universe with constant $c^2\rho = -p$, or to use our new notation, $w = -1$. To much satisfaction, $w = -1$ when plugged into the continuity equation is perfectly consistent with a constant ρ.

Finally, we should embrace our ignorance and admit there may be additional things in the universe we are not yet aware of.

- *Exotic components* – these unknown unknowns could potentially be anything at all, with any w. Cosmological scalar or vector fields, topological defects, phantom fluids, or a representation of a modification to gravity as a 'fictional' fluid might all fall into this category.

1.2.2 Simple solutions

Power laws

Despite not knowing the contents of our *actual* universe just yet, we can still look at how an imaginary universe evolves given a matter content. It can be difficult to get an analytic solution with more than one component, but if we approximate the universe as flat and dominated by a fluid that can be treated as having a single w, it is fairly simple. Friedmann's equation becomes

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho_0 a^{-3(1+w)}\,, \tag{1.50}$$

which solves to yield,

$$a \propto t^{2/3(1+w)}\,. \tag{1.51}$$

The larger the value of $|1 + w|$, then the slower the expansion. Normal matter ('dust') has $w = 0$, so $a \propto t^{2/3}$. For radiation ($w = 1/3$), $a \propto t^{1/2}$. For a fluid with $w = 1$ (referred to as 'stiff matter'), $a \propto t^{1/3}$.

de Sitter spacetime

The universe expands ever more rapidly with time as $w \to -1$, and everything goes wrong when $w = -1$ exactly. This is understandable as the solution is not valid for this case, so we return to solve Friedmann's equation for a constant ρ. That implies H is constant too. The solution is growth faster than a power-law could provide (hence why the form of the earlier solution refused to cooperate),

$$a \propto e^{Ht}\,. \tag{1.52}$$

Thus for a universe with constant ρ (effectively the case of a cosmological constant dominated universe), the universe expands exponentially. This sort of universe actually constitutes a named spacetime of its very own, and is called *de Sitter spacetime*.

We should also add that Friedmann's equation can be written

$$\frac{\Lambda}{3} \equiv H^2 = \frac{8\pi G}{3}\rho\,, \tag{1.53}$$

where $\Lambda/3$ is the constant cosmological constant term.

1.3 Observations from radiation dominated

1.3.1 Cosmic microwave background

Arguably the most important observation for cosmology is the discovery of a microwave background, of (nearly) constant temperature across the entire sky. This was one of the key predictions of the expanding universe (or 'hot big-bang') model. Knowing there is at least a little radiation in the universe, we can make some very general reasoning to see the origin of this microwave background.

Matter scales with the scale-factor as R^{-3} while radiation scales as R^{-4}. Inevitably, for some small enough R the universe will have been radiation dominated. Just as importantly, if the expansion is treated as adiabatic (a good approximation on large scales), then temperature also rises as R decreases. It is clear at some point in the past, the temperature will have been hot enough to ionise the atoms. The universe would have expanded and cooled from its high temperature, radiation dominated beginnings. Eventually the temperature (and, as importantly, the photon number density) would have dropped low enough for electrons to combine with nuclei to form neutral atoms. In a terrible choice of nomenclature, this is called recombination (despite there not being any 'combination' at a prior time). The photons were then free to travel through

space without being scattered further by charged particles. As this change took place so quickly it is reasonable to to consider the photons as being emitted from a thin spherical 'shell' that surrounds us, known as the surface of last scattering. Their expansion of the universe has since redshifted their wavelengths to the microwave, and this cosmic microwave background (CMB) was serendipitously discovered by Penzias and Wilson. [116]. Further, because the photons were so tightly coupled to the dissociated electrons at the time of last scattering, then the photons were in thermal equilibrium. Thus the spectrum was expected to be that of a blackbody, and this prediction was confirmed to be remarkably accurate [108].

At what redshift was the CMB emitted? The current radiation density is in the form of microwaves, with a temperature $T \approx 2.73K$. For adiabatic expansion, radiation has $\rho \propto T^4$ (recall the expression relating temperature and energy density for a black body). As $\rho \propto R^{-4}$ for a relativistic fluid, the temperature of radiation therefore scales as $T \propto 1/R$. A crude estimate for the temperature required to ionize the matter in the universe would be to set the temperature to require a photon to have the energy to ionize neutral hydrogen. Then $T = 13.6eV/k_{\mathrm{b}}$, where k_{b} is Boltzmann's constant (and we are relating energy to temperature using $E \sim k_{\mathrm{b}}T$). Unfortunately this neglects the fact that there are a huge number of photons in the universe compared to baryons, and that there will be plenty of photons in the long high energy tail of the distribution, capable of ionizing hydrogen. The temperature is significantly overestimated from the crude calculation.

A full analysis requires knowing the photon to baryon ratio, as well as solving the appropriate equation for the quasi-static statistical distribution of the photons and ionized hydrogen (the *Saha equation*). Rather than wading through this, we simply quote the result that recombination took place when the universe had cooled down to $T \approx 3000K$. Since $T \propto 1/R$, that means the universe was approximately one thousandth of the size, or $z_{\mathrm{rec}} \approx 1000$.

The universe is not completely isotropic and homogeneous. Even in the very early universe, some power spectrum of primordial perturbations existed, and were first observed with the COBE satellite [135] through temperature fluctuations in the CMB. These ripples are the seeds for structure (such as galaxies) to form in the universe. A huge array of information is available from these anisotropies, with the WMAP satellite continuing to make more detailed observations.

The CMB power spectrum is usually expanded in spherical harmonics, denoting the degree of the harmonic with the letter l. (This can be thought of as an analogue for

a Fourier transform on the surface of a sphere rather than a flat plane; the harmonic serves the same role as wavenumber, with smaller l indicating larger angular scales). On large scales the spectrum is dominated by the Sachs-Wolfe effect (caused by intrinsic perturbations in the background at last-scattering), while at some smaller angular scale Doppler and adiabatic effects become comparable. At small scales the adiabatic effects dominate. The end result is a peak in the spectra named the acoustic peak, at about $l \sim 100$. The height is dependent upon the amount of matter in the form of baryons Ω_b, while the angular position depends on Ω. The baryon fraction is constrained from WMAP 5-year data alone to be $100\Omega_b h^2 = 2.273 \pm 0.062$. In Λ dominated universes, the shift in the peak from varying Ω is almost negligible, however, which means separate observations must be included before flatness can be strongly constrained. However, highly curved models require a low value of H_0, which would be inconsistent with observation [54].

1.3.2 Nucleosynthesis

Proton-neutron ratio

Nuclear reactions occur at extremely large temperatures. The interior of stars is an example of this, with temperatures at tens of millions of Kelvin. However, when the universe was at such temperatures, the density was much lower than the interior of stars. It was only at much earlier times (with higher temperatures and densities) that the nuclear reactions became important in cosmology, with temperatures at $\sim 10^9$ K.

Thermodynamics can be applied to find a rough estimate of the proton to neutron ratio from these early-time nuclear reactions. The neutrons have a larger mass than the protons, giving that state a higher energy of Δmc^2. Boltzmann's hypothesis leads to a suppression factor of $e^{-\Delta mc^2/kT}$ for these higher energy states, so that

$$\frac{N_n}{N_p} = e^{-\Delta mc^2/kT} \approx e^{-1.5\left(10^{10}k/T\right)}. \tag{1.54}$$

As the temperature decreases, the ratio decreases exponentially. This would lead to a vanishingly small number of neutrons today, but for the fact that thermal equilibrium only lasts until the timescale for the nuclear reactions eventually (as the universe cools) becomes longer than the expansion timescale. The reactions occur so rarely that the neutron-proton ratio fixes at some specific value (a phenomena known as *freeze-out*). In practice, this turns out to be $N_n/N_p \approx 1/7$. Assuming all the neutrons end up in helium, this gives a prediction of the hydrogen to helium ratio, that is roughly consistent with observation (given the simplifications made in the calculation).

More detailed calculations of this process allow a calculation of the hydrogen to helium ratio. Unfortunately, it is only very weakly dependent upon the baryon density, making the helium abundance a measure Ω_b.

Other light element abundances

Nuclear fusion reactions are still energetically favourable all the way up to the production of iron. Still, by the time significant amounts of helium has been produced, the density and temperature of the universe has declined too much for any significant synthesis of heavier elements to occur.

The remaining elements left behind tend to be deuterium, ^{3}He, and extremely small fractions of other elements such as ^{7}Li and ^{7}Be, with fractions of $\sim 10^{-9}$ and $\sim 10^{-11}$ by mass, respectively.

1.4 Observations from matter domination

1.4.1 Supernovae

Recall the expansion of the universe leads to a gravitational redshifting of light emitted from distant sources. The redshift z is related to the wavelength of the emitted light and to the scale-factor R by

$$1 + z = \frac{\lambda_0}{\lambda} = \frac{R_0}{R} \tag{1.55}$$

where subscript zero as usual indicates the value of quantities as measured today.

By definition, the apparent magnitude of a source m is related to its absolute magnitude M via the equation

$$m - M = 5 \log_{10} \left(\frac{d_L}{\mathrm{M_{pc}}} \right) + 25 \,, \tag{1.56}$$

where the luminosity distance d_L is defined in terms of the absolute luminosity of a source L_s and the energy flux $\mathcal{F}$,

$$d_L^2 \equiv \frac{L_s}{4\pi \mathcal{F}} \,. \tag{1.57}$$

In the case of cosmology, the observed luminosity of the source, L_0, will be dimmed from its absolute value by two factors of $(1 + z)$ - once due to redshifting of photons, and once due to fewer photons being received, so that $L_s = L_0(1 + z)^2$. Including the geometrical effect of spatial curvature is easier if we rewrite the Robertson-Walker

metric using a new variable χ for the radial coordinate. It is related to r via,

$$\mathrm{d}\chi^2 = \frac{\mathrm{d}r^2}{1-kr^2} \tag{1.58}$$

and so,

$$\mathrm{d}\tau^2 = \mathrm{d}t^2 - R^2(t)\left(\mathrm{d}\chi^2 + f_k^2(\chi)(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right) \tag{1.59}$$

with

$$f_k(\chi) = \begin{cases} \sin\chi, & k=+1, \\ \chi, & k=0, \\ \sinh\chi, & k=-1\,. \end{cases} \tag{1.60}$$

Note that we have adopted units with $c = 1$ to simplify the mathematics. We will continue with this practice from now on, unless clarity requires otherwise (and it will be obvious when such cases arise).

Now it is much clearer how the spatial curvature affects the luminosity distance; the angular part of the metric has $f_k^2(\chi)$ instead of r^2. The flux will therefore be decreased by $f_k^2(\chi)$ instead of simply χ^2 (the flux is spread over a wider angle). So we have the following expression for the luminosity distance:

$$d_L = R_0 f_k(\chi_s)(1+z)\,. \tag{1.61}$$

The value of χ_s, the χ coordinate for the source, can be calculated by recalling photons travelling in the χ direction follow the null geodesic $d\tau^2 = 0 = -dt^2 + R^2(t)d\chi^2$. So integrating and using the chain-rule,

$$\chi_s = \int_0^{\chi_s} d\chi = \frac{1}{R_0}\int_0^z \frac{dz'}{H(z')}\,. \tag{1.62}$$

Thus, different cosmologies will result in different values of $m - M$, a quantity that can, with sufficient knowledge of the source, be observed directly.

A white dwarf is a star supported by the degeneracy pressure of the electrons within its interior (as fermions, they cannot occupy the same state, and so there is a cost for forcing them into a smaller and smaller space associated with promoting them to higher and higher energy levels). Trying to shrink such a star increases the density, and so the degeneracy pressure becomes larger. Naively, the degeneracy pressure can balance the gravitational force trying to shrink the star up to indefinite densities. After all, an electron trapped in a box of side length L due to its surrounding fellow fermions finds its energy increasing as $1/L^2$; if the star is shrunk slightly so that the same amount of matter resides in a smaller volume, the density ρ will change as $\rho \propto L^{-3}$. The energy

increases with the density as $E \propto \rho^{2/3}$. Meanwhile the gravitational energy associated with all these 'boxed' electrons will increase proportional to M^2/r in the volume of radius r. Again shrinking the radius slightly (with the same amount of matter within) only increases the energy as $\rho^{1/3}$.

However, the degeneracy pressure cannot increase indefinitely. Eventually the electrons become so energetic that relativistic effects come into play, and the degeneracy pressure does not increase as the non-relativistic analysis of a particle in a box might suggest. For a relativistic particle in a box, the energy only increases as $\rho^{1/3}$, the same as gravity (recall $\Delta p \sim \Delta x^{-1} \sim \rho^{1/3}$, and that $E \propto p$ for relativistic matter). The constant of proportionality now matters; for gravity it goes as M^2 while for the degeneracy pressure, the mass is irrelevant (only density is important). Thus for a *massive* enough star, eventually gravity will become stronger. This upper limit is known as the *Chandrasekhar limit*, beyond which a more massive star cannot be supported, and is about 1.4 solar masses.

When a white dwarf star exceeds the Chandrasekhar limit, the resulting explosion is referred to as a Type Ia supernova (SN Ia). This explosion is so bright it is visible over cosmologically significant distances. Because the processes governing the SN Ia are believed to be independent of redshift, there is a common absolute magnitude for all such events (or, to be more technical, a common relationship between the width of the light curve and the absolute magnitude, though the details need not concern us). This makes them excellent distance indicators, allowing for a measurement of H both locally and and this technique was exploited by Perlmutter et. al. in 1999 [117] to show that a flat universe ($\Omega_m + \Omega_\Lambda = 1$) required a significant amount of vacuum energy ($\sim 70\%$). In 2004, Riess et. al. [124] added further observations using the Hubble Space Telescope (HST) and showed that the universe transitioned from a period of deceleration to acceleration at a greater than 99% confidence level. From equation (1.45), it becomes clear that no universe without some component of energy density obeying an unusual equation of state ($w < -1/3$) can plausibly fit the data.

1.4.2 Non-relativistic matter

Galaxy clusters

Stars are the most easily observed form of matter in the universe (at least to the naked eye). They are almost exclusively found within galaxies. A small fraction of galaxies congregate in appropriately named *galaxy clusters*. These sometimes aggregate

themselves into much larger *super clusters*. Some measure of how much mass can be found in these large-scale structures can be found through measuring the light emitted from the stars, but unfortunately this is not very useful - most of the baryonic matter is actually found in the intergalactic medium in the form of hydrogen gas clouds.

Fortunately the hydrogen tends to be hot, having gained considerable energy from falling into the deep gravitational potential well of the cluster. This hot gas emits light in the form of X-rays, measurements of which can provide an estimate of the temperature - and so the kinetic energy of the gas. Assumed an equilibrium situation, the virial theorem can be applied to give an estimate of the size of the potential well, obviously related to the *total* mass of the cluster. Surprisingly, when this exercise is carried out, this does *not* come out as a ratio of unity; in fact the self gravity of the gas only accounts for a few percent of the total estimated mass in such clusters [154]. This suggests the mass in baryons (hydrogen gas) is only a tiny fraction of the mass in the universe.

Rotation curves

There is further evidence of a form of non-baryonic matter by looking at the dynamics of galaxies. When a galaxy - particularly a spiral galaxy - is viewed edge-on, the orbital motion of the stars about the centre of mass of the galaxy is detectable through the redshift. The spectra of stars (or even the 21cm emission line from neutral hydrogen) on one side of the galaxy will be slightly redshifted, due to its apparent motion away from us. Stars (or gas) on the other side of the galaxy will be in the opposite phase of the orbit, and so the light will be blue shifted. This shifting is simply related to the rotational velocity. Carrying this measurement out at different radii from the galactic centre allows for a measurement of how the velocity of the luminous matter varies with radius, a function known as the *rotation curve* or *velocity profile*.

The velocity profile should drop as $v \propto r^{-1/2}$ once at a sufficient radius that the interior mass is approximately constant. Surprisingly, this does not happen. Instead, even as the *luminosity* dies away at larger radii, the rotation curve remains flat. If we are not willing to discard the established theory of Newtonian gravity, the only other explanation is that there is some form of non-luminous matter that extends far beyond – and rapidly dominates the contribution from – the luminous stars and gas.

Gravitational Lensing

It is fortunate there is a way to observe the gravitational mass directly. We are able to do this by observing the way in which light propagates through the universe. The gravitational fields of collapsed structures deflect the paths of light-rays. This effect is called *gravitational lensing*, as the mass distribution serves in much the same way as a more familiar lens made of glass does. Some knowledge of the lens (mass distribution) can be be inferred from the way the light is deflected between the source and our telescopes.

In the case of *strong* lensing, the effect is particularly pronounced. With near perfect alignment of source and lens, this can occur to the extent that multiple images of the same source can appear on the sky (the result of the curvature of spacetime being great enough to curve light around in multiple ways from a distant source to our telescopes). These can sometimes form dramatic arcs, or even an entire ring of light, known as an Einstein ring.

Usually spacetime is not so curved, and the alignment not so favourable. In this case the light is only mildly deflected by the intervening mass distribution. This is the domain of *weak* gravitational lensing. It then becomes a statistical exercise in examining the distortions to the shapes of background objects, to piece together or 'map' the underlying mass distribution [141].

The conclusions from gravitational lensing (e.g. [74, 41, 22]) only serve to reinforce conclusions about mass-to-light ratios being far higher than can be understood from baryons alone.

Dark Matter

Our theories of gravity relates the matter distribution to the gravitational forces. When the equations appear inconsistent, one or the other is mistaken.

There are historical examples of both occurrences. The discovery of Neptune was predated by its predicted existence, when irregularities in the orbits of other bodies in the solar system could not be explained by Newtonian gravity. Meanwhile the perihelion precession of Mercury was explained by modifying gravity, and supplanting Newton's equations with the full equations of Einstein's general relativity.

Cosmology is principally an application of gravitational physics. When consistency is not forthcoming, the dilemma of modifying gravity or introducing a new form of matter (or properly, in general relativity, energy-momentum) becomes apparent. The

cosmological constant straddles the boundary between these two options; depending on physical interpretation, it could function as either a smooth, uniform form of energy momentum, or the simplest modification to the field equations.

The peculiar galactic dynamics, as well as the surprisingly large mass in galaxy clusters, gives cosmologists just this dilemma. The widely accepted choice is that our gravitational theories are correct, and there exists a large, unobserved matter component that balances the equations.

As this dark matter forms the bulk of the matter in he universe, its nature has a significant impact on the formation of large scale structure. If this component easily undergoes gravitational collapse, the baryonic matter will cluster more easily due to the influence of the gravitational wells. Obviously in the contrary case, structure formation will be not be so great if dark matter does not cluster.

A conceptually simple way to test what sort of dark matter is appropriate is to examine the theoretical predictions of structure formation for different types of dark matter, and then compare with the observed large scale distribution of matter. This has essentially been done, using modern supercomputing power to solve Newton's Laws in a comoving reference frame. These simulations are usually done using particulate dark matter, with hydrodynamic modelling included to simulate the gaseous baryons.

A distinction can be made between *hot* and *cold* dark matter. Hot dark matter behaves as a relativistic fluid, while cold dark matter behaves as a non-relativistic one. A middle ground is quite sensibly referred to as *warm* dark matter. The analytic details of structure formation at a linear level are delayed until a later chapter, but the results of the simulations confirm the intuitive result. Cold dark matter, having a lower velocity, clusters more easily, allowing for faster structure formation.

In fact, comparison to galaxy surveys [153, 11] shows cold dark matter to be a good match to the way structure has formed. The main caveat to this objection occurs at small scales; the simulations predict an overabundance of small dark matter halos that have not been observed. The usual answer to this is both that the small scale physics of the baryons is poorly modelled, and crucially only the baryons are observed. Small dark matter halos with a small amount (or a poor mass-to-light ratio) of baryons are difficult to detect. Indeed, many subhalos have been observed that were originally unknown, alleviating the problem somewhat.

MOND

The most common alternative to cold dark matter is a phenomenological model called Modified Newtonian Dynamics (MOND). The core idea behind this proposal is that either Newtonian dynamics or the gravitational force changes as a function of the gravitational acceleration. The actual function is left unspecified, but in the limit of accelerations much larger than some characteristic accelerations a_0, Newtonian physics is recovered. At much smaller accelerations (such as those at the distance of galactic radii), the gravitational force upon an object no longer depends upon m, but ma_0/a, with a as the object's acceleration. This naturally leads to flat rotation curves, and led Milgrom to propose the original idea [109].

MOND has had some success at the galactic level, though any applications to cosmology require use of the relativistic version *Tensor-Vector-Scalar gravity* or (TeVeS), which has significantly more freedom than MOND (which is essentially fixed by a single free parameter, a_0). However, one particular observation has left MOND in great difficulty.

The 'bullet cluster' observation [41] refers to a pair of colliding galaxy clusters, with the smaller one having passed through the centre of the larger one. This sounds innocuous enough, but the implications are important for distinguishing MOND from dark matter. As the two galaxy clusters collide, the gas (baryons) in the clusters becomes heated. The X-ray emission from the hot baryons is easily detectable, with the interaction between the baryons causing them to stall at the point of collision.

Meanwhile gravitational lensing is sourced by the total mass distribution. The major dark matter halos are not aligned with the baryons; in particular, the smaller 'bullet' cluster has a large amount of dark matter ahead of the baryonic gas. This is easily understandable with dark matter being explained by only a weakly interacting matter component; for MOND it proves very difficult.

1.4.3 Baryon acoustic oscillations

Baryon acoustic oscillations (BAOs) are a curious phenomena resulting from the tight coupling between photons and baryons in the early (pre-CMB) universe.

Overdense regions in the early universe also have a large pressure, due to the hot relativistic particles that are coupled to the baryons. An initial overdensity results in a spherical sound wave, with the perturbation in gas and photons carried outwards away from the initial source. (The perturbations are small enough that the combination of

many of these overdensities can just be summed linearly.) The dark matter remains at the centre, uncoupled to the other components of the system.

As the universe cools, photons recombination occurs and the photons are able to travel long distances. They do not scatter efficiently from the newly formed neutral atoms, and so the coupling between baryons and photons is no longer very strong. The radiation travels unimpeded (until it reaches our telescopes) while there remains an overdensity at about 150Mpc away from the central region; this being the distance the sound wave could travel before recombination. With the sound wave now stalled due to decoupling, the baryons and dark matter are only influenced by their mutual gravitational attraction. Due to the large amounts of dark matter, the baryon overdensity ends up correlated with the dark matter overdensity once again. Despite this, there remains a small residual peak in the correlation function of galaxies at this characteristic distance of 150Mpc, or (viewed in the power spectrum, a series of acoustic oscillations, hence the name) [18].

The first claim of detection was only relatively recent [42, 58]. Not only is this pleasantly consistent with our knowledge of cosmology but the characteristic scale of this effect can be measured at multiple redshifts, i.e. it serves as a statistical *standard ruler*, and potentially one that can be observed at a range of redshifts. This allows for test of spatial flatness, as well as a measurement of $H(z)$. Combined with the CMB [73], we now know the universe to be extremely close (if not exactly) flat.

1.5 ΛCDM cosmology

Let us now summarise, and put all of this together. From the CMB and BAOs, the universe is found to be flat, so $\Omega \approx 1$. These measurements also demand a very low baryon fraction ($\Omega_b \approx 0.04$). Comparing light element abundances with nucleosynthesis predictions also requires a low baryon fraction. Cosmological observations therefore suggest baryons only constitute a tiny fraction of the total amount of matter in the universe. This is surprising claim, but the same conclusion has been drawn from galaxy rotation curves. The dynamics are only consistent if there is a large, non-luminous 'halo' of matter that dominates the mass profile at large galaxy radii. Finally, X-ray observations of galaxy clusters suggests a gravitational potential well far deeper than can be explained by the stars and gas of the cluster; further evidence for some 'dark matter'.

A natural assumption might then be that the amount of dark matter is sufficient

to make the universe flat (so the total dark matter and baryon content $\Omega_m \approx 1$). Unfortunately, not only is this far too much dark matter to match the observations, but such a universe does not accelerate. Supernova data therefore rules this scenario out.

A third component is needed to generate the accelerated expansion. This is the dark energy. There remains a great deal of uncertainty as to its nature, but the simplest option is a cosmological constant. The best fit model, given all of the above data, has $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$ [90].

This model of Λ and cold dark matter (ΛCDM) fits and explains the data remarkably well. The cold dark matter paradigm predicts the large scale structures we observe, and although simulations tend to overproduce the smaller halos [111], it is widely believed this is a result of poor simulation of baryon physics [28], and the fact many of the small dark matter halos are so far undetected.

The cosmological constant is the simplest modification to gravity, and so far all observations are consistent with this explanation. On the other hand, the observations are consistent with a host of alternative models, also. The driving arguments separating models of dark energy are, at this stage, mostly theoretical in nature. Troubling theoretical issues regarding the magnitude of Λ have motivated a plethora of other scenarios, but so far none has appeared compelling enough to abandon the simplicity of the standard cosmological model. Many of the alternatives have theoretical difficulties of their own, as we shall see in the following chapters.

Chapter 2

Field Theory and Cosmology

2.1 Introducing field theories

Before embarking on listing and studying the huge plethora of dark energy models that now persist, it is worth spending a little time recalling the basics of classical, relativistic field theory. Many of the basic mathematical concepts (principally tensors) have been introduced in the previous chapter.

Many cosmological models involve classical fields. In fact, like all other classical systems, this framework is only an approximation to an underlying quantum description of the world. Modern particle physics is in fact based upon the quantised field theory framework. For now we will satisfy ourselves that the classical description should provide a good approximation in the appropriate domain.

The most familiar example of a classical field theory is that of electromagnetism, which uses the four-vector A_μ as the electromagnetic potential. Cosmological models usually make use of far simpler scalar fields, usually denoted with the symbol ϕ. Part of this section will be looking at what makes these fields so attractive for cosmological models.

2.2 The Lagrangian

Rather than writing down field equations directly, it is commonplace to work from a *Lagrangian density* $\mathcal{L}$. Often this is just referred to (slightly confusingly) as the Lagrangian, and this common shorthand shall certainly be used here, when the context makes the full meaning obvious.

The Lagrangian density $\mathcal{L}$ is related to the familiar Lagrangian L, which is used in the action formalism of point particles in classical mechanics. Recall that particles follow paths that minimise the action $S \equiv \int L \, \mathrm{d}t$. If q is the generalised coordinate of the particle, and $\dot{q} \equiv \mathrm{d}q/\mathrm{d}t$ is velocity, then

$$\delta S \equiv \delta \int_{t_1}^{t_2} L(q, \dot{q}) \mathrm{d}t = 0 \, . \tag{2.1}$$

Variational calculus needs to be applied to solve this problem. If small variations $\delta q(t)$ to the path $q(t)$ are considered, then to first order:

$$\delta L = \frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q} \, , \tag{2.2}$$

with $\delta \dot{q} = (\mathrm{d}/\mathrm{d}t) \, \delta q$.

Substituting this into the variation of the action, integrating by parts, and imposing the condition that the end-points of the path remain unchanged ($\delta q(t_1) = \delta q(t_2) = 0$), we find the *Euler-Lagrange Equation*,

$$\frac{\partial L}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial L}{\partial \dot{q}} \right) = 0 \, . \tag{2.3}$$

For multiple generalised coordinates, we might write down a Lagrangian for each coordinate:

$$L = L_x + L_y + L_z = \sum_i L_i \, , \tag{2.4}$$

where the sum has been generalised to any number of degrees of freedom.

2.3 The Lagrangian density

A field ϕ is more complicated than a single coordinate $q(t)$; one way of viewing the complexity is going from a single degree of freedom q at each point in time, to a degree of freedom $\phi(x)$ at each point in spacetime (labelled here generically as x, no matter the number of dimensions). How do we handle the infinite number of degrees of freedom we have found ourselves with, due to the continuity of spacetime?

In this case, the Lagrangian becomes not a sum but an integral over a *Lagrangian density* $\mathcal{L}$ which depends, in the general sense, on the field and its derivatives with respect to space and time, denoted $\partial_\mu \phi$.

$$L = \int \mathrm{d}x^{d-1}\, \mathcal{L}\left(\phi, \partial_\mu \phi\right) . \tag{2.5}$$

Here d indicates the number of dimensions; typically four (one of time, three of space). The action already looks a lot more relativistic this way;

$$S = \int \mathrm{d}t \int \mathrm{d}^{d-1}x\, \mathcal{L}\left(\phi, \partial_\mu \phi\right) = \int \mathrm{d}^{d}x\, \mathcal{L}\left(\phi, \partial_\mu \phi\right) . \tag{2.6}$$

where we have shifted the definition of x to include time as well as space, to make it clear how equally all the dimensions are treated.

Considering small variations $\delta\phi(x,t)$ of the field, and assuming they vanish on the integration boundary (which removes surface terms), results in the Euler-Lagrange equation for ϕ,

$$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu}\left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right) = 0 . \tag{2.7}$$

For general relativity, only a small tweak is needed - rather than integrating over $\mathrm{d}x^d$, instead one must integrate over $\sqrt{-g(x)}\mathrm{d}x^d$. This is simply to ensure the action S remains a scalar; changes to the coordinate system $\mathrm{d}x^d$ are now compensated for by the changes to the determinant of the metric $g(x)$, ensuring $\sqrt{-g(x)}\mathrm{d}^d x$ is an invariant.

2.4 Field equations

This has been very general. It would be useful to examine some specific Lagrangian densities, and a specified spacetime, and see what the field equations look like.

2.4.1 Canonical scalar field

The Lagrangian density for a scalar field is typically separated into a part containing derivatives of the field, minus a part without derivatives (depending only upon the field). The former is called the kinetic terms, the latter the potential. For a 'canonical' scalar field ϕ, the kinetic part is $\frac{1}{2}\partial_\mu \phi\, \partial^\mu \phi$. The Lagrangian density can then be written

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi\right)^2 - V(\phi) . \tag{2.8}$$

Here $\left(\partial_\mu \phi\right)^2$ has been used as short-hand for the scalar product, $\partial_\mu \phi\, \partial^\mu \phi$, which looks clumsy when written so formally.

The most simple of potentials $V(\phi)$ is just quadratic in the field,

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi\right)^2 - \frac{1}{2}m^2\phi\,. \tag{2.9}$$

The parameter m is often referred to as the mass parameter.

Lagrangians describing fields that have only quadratic potentials (including the massless $m = 0$ case) are referred to as *free* field theories. Other terms in the potential describe self-interactions or interactions with other fields.

The field equations take on the following form:

$$\partial^\mu \partial_\mu \phi + \frac{\partial V}{\partial_\phi} = 0. \tag{2.10}$$

This becomes the well known Klein-Gordon equation when a quadratic potential is used

$$\left(\partial^\mu \partial_\mu + m^2\right)\phi = 0\,. \tag{2.11}$$

2.4.2 Electromagnetism

As an example of an action for a vector field, we can also write an action for electromagnetism. To do so we define the *field tensor* $F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu$, where A^μ is the vector potential. Maxwell's equations are then $\partial_\nu F^{\mu\nu} = -\mu_0 J^\mu$ (μ_0 being the permeability of free space, and J^μ being the current four-vector).

The Lagrangian density for this equation is,

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \mu_0 J^\mu A_\mu\,. \tag{2.12}$$

2.4.3 Gravity

Remarkably, there exists an action principle for gravity also. Einstein's field equations are generated from the *Einstein-Hilbert Lagrangian*:

$$\mathcal{L} = -\frac{1}{16\pi G}\left(R + 2\Lambda\right)\sqrt{-g} + \mathcal{L}_{\text{matter}}\,. \tag{2.13}$$

Here $\mathcal{L}_{\text{matter}}$ is the Lagrangian density for the matter.

2.4.4 Noether's Theorem

The Lagrangian formulation of classical mechanics provides a useful insight into conservation laws. Recall Euler's equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0\,. \tag{2.14}$$

Here the generalised coordinates are q_i. If we define generalised momenta as $p_i = \partial\mathcal{L}/\partial\dot{q}_i$, then we can rearrange Euler's equation,

$$\frac{\partial\mathcal{L}}{\partial q_i} = \dot{p}_i \,. \tag{2.15}$$

When the Lagrangian is invariant under a coordinate, the momentum for that coordinate is conserved.

These invariances manifest themselves as conservation laws in relativistic field theory also, though rather than a conserved quantity there is a conserved *current*. To see this requires a consideration of certain continuous transformations of the field(s) ϕ, which can be written in infinitesimal form as

$$\phi(x) \to \phi'(x) = \phi(x) + \alpha\Delta\phi(x) \,, \tag{2.16}$$

taking α as an infinitesimal parameter and $\Delta\phi(x)$ some change in the field configuration. Such a transformation is called a symmetry if it leaves the equations of motion invariant. This is not only the case if the action remains unchanged by transformations as (2.16), but the action may even be changed by a surface term and leaves the equations of motion unchanged (recall the surface terms vanish in the derivation). A surface term in the action is produced by a four-divergence in the Lagrangian. Therefore if we are demanding invariant equations of motion under this sort of transformation, the Lagrangian is invariant under (2.16) up to a four-divergence:

$$\mathcal{L}(x) \to \mathcal{L}(x) + \alpha\partial_\mu\mathcal{J}^\mu(x) \,, \tag{2.17}$$

for some four vector $\mathcal{J}^\mu$.

We can find out how $\mathcal{J}^\mu$ is related to deformations in the field $\Delta\phi$ by the above change $\Delta\mathcal{L}$ to the result from varying the fields:

$$\begin{aligned} \alpha\Delta\mathcal{L} &= \frac{\partial\mathcal{L}}{\partial\phi}(\alpha\Delta\phi) + \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\partial_\mu(\alpha\Delta\phi) && (2.18)\\ &= \alpha\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\Delta\phi\right) + \alpha\left[\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\right]\Delta\phi \,. && (2.19)\end{aligned}$$

The rearrangement serves a purpose; the second term vanishes by the Euler-Lagrange equation. Then we relate $\Delta\mathcal{L}$ calculated in these two different approaches;

$$\alpha\Delta\mathcal{L} = \alpha\partial_\mu\mathcal{J}^\mu = \alpha\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\Delta\phi\right) \,, \tag{2.20}$$

or

$$0 = \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\Delta\phi - \mathcal{J}^\mu\right) \,. \tag{2.21}$$

Thus if we make the definition:

$$j^\mu(x) \equiv \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \Delta\phi - \mathcal{J}^\mu \,, \tag{2.22}$$

then we know it a conserved four-current,

$$\partial_\mu j^\mu(x) = 0 \,. \tag{2.23}$$

This is a powerful result: every continuous symmetry (by which is meant an invariance of the Lagrangian under a transformation of the form we have been discussing) leads to a conserved current.

To see a simple example in practice, consider $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2$. The transformation $\phi \to \phi + \alpha$, where α is a constant, leaves $\mathcal{L}$ unchanged. From this we know $\mathcal{J}$ is zero, thus the current $j^\mu = \partial^\mu \phi$ is conserved.

Now we finally come to the practical point of all of this mathematics. Noether's theorem can be applied to spacetime transformations, such as translations and rotations. Obviously the action should not be changed by such a transformation. To see the consequences, we will need to rewrite an infinitesimal translation,

$$x^\mu \to x^\mu - a^\mu \,, \tag{2.24}$$

as a transformation of the field configuration,

$$\phi(x) \to \phi(x + a) = \phi(x) + a^\mu \partial_\mu \phi(x) \,. \tag{2.25}$$

As the Lagrangian is a scalar, it transforms in the same way,

$$\mathcal{L} \to \mathcal{L} + a^\mu \partial_\mu \mathcal{L} \,. \tag{2.26}$$

We would like to be able to deduce $\mathcal{J}^\mu$ from this, from (2.21), so we need the second term in the form $\alpha \partial_\mu J^\mu$. We can write,

$$a^\mu \partial_\nu \mathcal{L} = a^\nu \partial_\mu \left(\delta^\mu{}_\nu \mathcal{L} \right) \,. \tag{2.27}$$

We can deduce $\mathcal{J}^\mu = \delta^\mu{}_\nu$. Actually there are four separate $\mathcal{J}^\mu$ here, one for each ν. There are then four separately conserved currents,

$$T^\mu{}_\nu \equiv \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi - \mathcal{L} \delta^\mu{}_\nu \,. \tag{2.28}$$

This is the energy-momentum tensor for the field; just as energy is conserved in classical mechanics due to time invariance of the action, so energy-momentum is conserved for relativistic fields due to the invariance of the action to translations in spacetime.

We can compare the energy-momentum tensor $T^{\mu}{}_{\nu}$ for the scalar field with that of a perfect fluid. Recall in special relativity this is just $\mathrm{diag}(\rho, -P, -P, -P)$. Denoting the momentum density conjugate of the field as

$$\pi(x) \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}(x)} = \dot{\phi}\,. \tag{2.29}$$

We can then read off the energy-density for a scalar field:

$$\rho = T^{00} = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\left(\nabla\phi\right)^2 + V\,. \tag{2.30}$$

We might interpret these terms as the energy cost of motion of the field, the cost of gradients in the field, and the cost of having the field take non-zero values.

The pressure can also be calculated.

$$P = T^{ii} = \frac{1}{2}\dot{\phi}^2 - V(\phi) + (\partial^i\phi)^2 - \frac{1}{2}(\nabla\phi)^2\,. \tag{2.31}$$

(Note none of these repeated indices are summed!) There is an annoying $(\partial^i\phi)^2$ term ruining the neat result. But since this is a perfect fluid, $P = T^{11} = T^{22} = T^{33}$, and so

$$P = \frac{1}{3}\sum_i T^{ii}\,. \tag{2.32}$$

If we do this sum, we can group these pesky terms up:

$$\sum_i (\partial^i\phi)^2 = (\nabla\phi)^2\ . \tag{2.33}$$

Thus,

$$P = \frac{1}{2}\dot{\phi}^2 - V(\phi) - \frac{1}{6}(\nabla\phi)^2\,. \tag{2.34}$$

2.5 Scalar fields in cosmology

Scalar fields are used frequently in cosmological models. Most of the above follows through, with the exception that the metric needed to be used is FLRW spacetime, rather than Minkowski.

2.5.1 Equation of motion

Applying the Euler-Lagrange equations with this metric leads to a different equation of motion. Space and time derivatives have been separated out,

$$\ddot{\phi} + 3H\dot{\phi} - \frac{1}{a^2}\nabla^2\phi = -\frac{\partial V}{\partial\phi}\,. \tag{2.35}$$

Since cosmological fluids should be *smooth*, the gradient term should be able to be neglected; particularly as the effect of the expansion is to reduce the potency of this term anyway.

Without the gradient term, this equation looks very similar to that of a classical particle moving along a single dimension ϕ, with an a potential energy V. The term involving Hubble's parameter acts just as friction, though obviously the situation is more complicated as H is time dependant and will depend in some way upon ϕ and $\dot{\phi}$, making the situation potentially highly non-linear. Fortunately there are usually some simple solutions.

Equation of state

A useful characteristic of cosmological fluids is the equation of state parameter, $w \equiv P/\rho$. For the perfect fluids considered earlier, this parameter was a constant. For a scalar field, we can use our earlier result (and neglecting gradient terms for a homogeneous universe), to write down:

$$w = P/\rho = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}\,. \tag{2.36}$$

The equation of state therefore always takes a value $-1 \leq w \leq 1$, and where it lies in this range depends upon the ratio of the kinetic to potential energy. In the case that the kinetic energy dominates, $\dot{\phi}^2 \gg V(\phi)$, then $w \approx 1$ and the scalar field behaves as stiff matter, with $\rho \propto a^{-6}$. On the other hand, if $\dot{\phi}^2 \ll V(\phi)$, then $w \approx -1$ and the scalar field instead behaves in a very similar fashion to a cosmological constant. Scalar fields can therefore produce a range of different behaviours. It is perhaps unsurprising then that they are so often used in cosmological models.

2.5.2 Problems with the standard model

While the standard model is remarkably successful at explaining a variety of independent observations over a wide range in redshift, the initial conditions of the universe remain particularly troubling. The particular difficulties can be expressed as a number of 'problems'.

Horizon problem

The standard cosmological picture has a particle horizon, with a comoving radius of

$$r_H = \int_0^t \frac{c\,dt'}{R(t')}. \tag{2.37}$$

If $R(t) \propto t^n$, then for $n < 1$, $d_H(t)$ is proportional to t^{1-n}. As the universe gets older, more of the universe is in causal contact. As the universe gets younger, a smaller and smaller part is in causal contact. This becomes a troubling point when considering the smoothness of the CMB, which is uniform to one part in 10^4 even when opposite points on the sky are compared. The particle horizon at that redshift, on the other hand, corresponds to a proper distance $D_H = R_0 r_H$ of a mere ~ 100 Mpc, which subtends an angle of about 1 degree on the sky. There remains no reason for this large scale homogeneity or isotropy, and the model that we have concluded from this assumption seems to preclude there ever being a physical process that could cause it.

Flatness problem

Observations have found that $\Omega \approx 1$. But such a universe is clearly very strange, as $\Omega(t)$ has a natural tendency to deviate from any value close to, but not precisely unity, i.e. such a condition is unstable. This suggests that Ω must have been extraordinarily fine-tuned close to unity at early times, in order to give a value consistent with observations today. To be more specific (from [89]),

$$\begin{aligned} |\Omega(10^{-43}\text{sec}) - 1| &\leq \mathcal{O}(10^{-60}), \\ |\Omega(1\text{sec}) - 1| &\leq \mathcal{O}(10^{-16}). \end{aligned} \tag{2.38}$$

This fine-tuning is clearly a deficit to the model.

Origin of initial density perturbations

The universe is not entirely homogeneous. The evidence is obvious in the present epoch, and even at the furthest redshifts observed, where the CMB anisotropies show a nearly scale invariant power spectrum of density fluctuations. The model is silent on the origin of these perturbations.

2.5.3 Introducing inflation

Having seen all these difficulties, it would be nice to have an underlying explanation for all of them at once. There is indeed such an explanation, and it is known as cosmic inflation. While inflation is by no means as well established as the 'standard model' cosmology, it is important to discuss due to the similarities between the behaviour of an inflationary universe and today's accelerated expansion. Naturally this has led to many dark energy models being inspired by cosmic inflation, making a discussion of

inflation even more important when discussing dark energy models. Along with this, inflation has had a surprising amount of success beyond its original explanatory goals, making it perhaps the most standard non-standard addition to modern cosmological models.

The basic idea behind inflation [97] is that there was an epoch when the universe expanded as if dominated by a cosmological constant, expanding exponentially. Obviously if *actually* dominated by a cosmological constant, there would be no returning to the radiation and then matter dominated universe of standard cosmology (the matter components redshift away with the expansion very rapidly, while the vacuum energy cannot be diluted by its very nature), so some mechanism is needed to mimic this behaviour, and eventually transition to a more familiar universe. This brings us to the namesake of this chapter; a scalar field.

Consider a universe dominated by the energy of a smooth, scalar field ϕ with some potential V. It obeys, as we have seen, the equation of motion

$$\ddot{\phi} = 3H\dot{\phi} + V'(\phi) = 0\,, \tag{2.39}$$

where we have neglected spatial derivatives (we have assumed the field is smooth). The potential we have left as a free function.

Now consider the case where the gradient of the potential is sufficiently small that the field 'rolls[1]' only slowly down a potential, such that the time Δt for ϕ to roll to its minimum is long compared to the expansion timescale, so $H\Delta t \gg 1$. Along with this, we imagine that the field rolls so slowly that the kinetic energy of the field is much less than the potential energy, $\dot{\phi}^2/2 \ll V(\phi)$. We have already seen that in such a case, a scalar field has an equation of state parameter $w \approx 1$, acting much like a cosmological constant. The scalar-field is then capable of driving a period of exponential expansion (de Sitter phase).

We can immediately see how this resolves the horizon problem. With this 'faster than light' expansion, then if $R(t) \propto t^n$ with $n > 1$, the integral for the comoving horizon distance actually becomes larger at smaller time. A period of such rapid expansion means that at earlier times, ever more of the universe was in causal contact.

This also serves to solve the flatness problem. Consider the Friedmann equation,

$$\dot{a}^2 = \frac{8\pi G\rho a^2}{3} - kc^2\,. \tag{2.40}$$

[1]Much of the nomenclature in this subkect comes from the fact that the equation of motion for a scalar field is identical to that of a ball rolling down a slope, with potential energy V and friction $3H$.

The first term must be hugely dominant over the second at the earliest times, in order to produce the flatness we observe today. An inflationary period where ρa^2 actually increases as the universe expands (unlike matter or radiation dominated phases) can obviously bring such a condition about, provided the universe inflates for sufficiently long.

The amount of inflation is usually written as the number N of *e-foldings*. If the universe expands from an initial scale-factor a_1 to a final scale-factor a_2, then $N = \ln(a_2/a_1)$. Both the flatness problem and the horizon problem require the same number of e-foldings to be resolved. Assuming that at temperatures of 10^{15} GeV (the typical energy scale where *grand unified theories* (GUTs) of particle physics are believed to be relevant - see reference) the universe was not fine-tuned, inflation resolves both of the problems with about $N \geq 60$.

Unlike a true de Sitter universe, this inflationary epoch need not last forever. Eventually ϕ reaches the minimum of its potential, resulting in the kinetic energy of the field no-longer being sub-dominant to the potential. The exponential expansion ends, and as the field oscillates about its potential minimum, it decays into lighter particles it is coupled to, restoring matter and radiation to the universe.

2.5.4 Inflation dynamics

Neglecting spatial derivatives, the equation of motion for the field ϕ is

$$\ddot{\phi} + 3H\dot{\phi} + dV/d\phi = 0\,. \tag{2.41}$$

The neglect of the $\nabla^2\phi$ term is easily justified, as the value of $\nabla\phi = \nabla_{\text{comoving}}\,\phi/a$ is rapidly diminished when a expands exponentially. There is usually a further simplification made by invoking the *slow-roll approximation*, whereby the acceleration $\ddot{\phi}$ is treated as negligible when compared to $3H\dot{\phi}$ and $dV/d\phi$. These conditions are actually a prerequisite for inflation to take place; the value of w is only close to -1 if ϕ has only small spatial gradients, and evolves only slowly with time.

Following this, the dynamical equation for ϕ is then given by the *slow-roll equation*,

$$3H\dot{\phi} = -dV/d\phi\,. \tag{2.42}$$

The friction term is balanced by the acceleration produced by the slope of the potential.

The condition $\dot{\phi}^2/2 \ll V(\phi)$ can be combined with the slow-roll equation to give a consistency condition,

$$\epsilon \equiv \frac{m_{\text{pl}}^2}{16\pi}\left(\frac{V'}{V}\right)^2 \ll 1\,. \tag{2.43}$$

Here we are using the Planck mass, which (in units where $c = \hbar = 1$, $m_{\rm pl} = \sqrt{G}$, while primes indicate partial derivatives with respect to ϕ.

Differentiating ϵ and using the slow-roll equation and $\dot{\phi}^2/2 \ll V(\phi)$ once again produces another condition,

$$\eta \equiv \frac{m_{\rm p}^2}{8\pi}\left(\frac{V''}{V}\right) \ll 1\,. \tag{2.44}$$

These two parameters are known as the *slow-roll parameters*, and are used to provide a consistency check on the slow-roll approximation.

Since the energy of the field is dominated by the potential energy,

$$H^2 = \frac{8\pi}{3m_{\rm p}^2}V(\phi)\,. \tag{2.45}$$

Combined with the slow-roll equation, it is easy to calculate the amount of growth the universe experiences

$$\ln(a_1/a_2) = N(\phi_1 \to \phi_2) = \int_{t_1}^{t_2} H{\rm d}t = -\frac{8\pi}{m_{\rm pl}^2}\int_{\phi_1}^{\phi_2}\frac{V(\phi)}{V'(\phi)}{\rm d}\phi\,. \tag{2.46}$$

A rough approximation for smooth potentials gives $V' \sim V/\phi$, so $N \sim (\phi_{\rm start}/m_{\rm pl})^2$. Inflation is successful if the initial value of ϕ is much larger than $m_{\rm pl}$. By the same reasoning, this is the same criterion to ensure the slow-roll parameters are both small. We can see almost any model of inflation that has a sufficiently flat potential as to allow inflation will have little difficulty producing enough e-foldings to resolve the cosmological problems.

Two of the most popular models of inflation are the quadratic $m^2\phi^2$ and quartic $\lambda\phi^4$ potentials. Higher order polynomials are possible but generally neglected due to difficulties when including quantum field theory (see reference chapter). For these models, the slow-roll approximation is extremely successful. The values of m and λ are at this stage completely unconstrained; ideally there would be some fundamental particle physics model that explains the nature of ϕ and predicts these values. In the absence of such a theory, observational constraints must instead be used. Fortunately these parameters are related to the primordial density fluctuations, imprinted upon the CMB, and so can indeed be constrained.

2.5.5 Reheating

Before discussing the origin of primordial density fluctuations, we should turn to the process by which the universe is returned to one filled with matter and radiation, rather than dominated by a scalar field.

Once ϕ approaches its minimum, the slow-roll parameters eventually become large enough that the slow-roll approximation is invalid, and inflation ceases. The change in behaviour is particularly dramatic when ϕ begins oscillating about this minimum on a timescale much shorter than the expansion timescale. Its energy then begins to be deposited in those of other fields it is coupled to (such as matter and radiation), 'reheating' the universe [86].

If the oscillations are large enough, the decay of ϕ into matter fields ψ_i first begins with preheating. With a coupling g between ϕ (with mass m_ψ and ψ, and writing

$$\psi = a^{-3/2}\sigma\,, \tag{2.47}$$

then the equation of motion for the (rescaled) matter field is

$$\ddot{\sigma} + \left(\epsilon^2 + g^2\phi^2 - \frac{3}{4}H^2 - \frac{3}{2}\frac{\ddot{a}}{a}\right)\sigma = 0\,, \tag{2.48}$$

where $\epsilon^2 = \frac{k^2}{a^2} + m_\psi^2$ is the frequency squared of the ψ oscillations (interpreted as the energy squared of ψ quanta).

The last two terms are sufficiently small to be ignored. The resulting equation describes an oscillator with a time dependent frequency. The effect is a parametric resonance; the inflaton decays into certain k-modes in ψ with a much higher probability than others. This can be interpreted as accounting for the effects of Bose statistics: once ψ particles are already produced, the decay rate may be greatly enhanced. The full analysis is complicated and will not be covered here (see [86, 70] for a full treatment).

Unless this process is incredibly efficient (transferring all the energy almost immediately), the particles produced from preheating are redshifted away by the expansion. As the oscillations of ϕ becomes small, particle production is dominated by a different sort of mechanism. The field ϕ continues to make full its ocillations about the minimum. In this case $\dot{\phi}^2$ can be replaced by its average over an oscillation cycle. When the potential is quadratic in the field, the equation of motion is just one of a damped harmonic oscillator, so

$$\langle\dot{\phi}^2\rangle = \rho_\phi\,. \tag{2.49}$$

Also recall that for a simple harmonic oscillator $\langle V\rangle = \langle\dot{\phi}^2\rangle/2 = \rho_\phi/2$, and so the pressure $P_\phi = \dot{\phi}^2/2 - V(\phi)$ vanishes.

While this oscillating behaviour takes place ϕ will decay into lighter particles it is coupled to. While undergoing this behaviour, a good approximation is to introduce an additional damping term $\Gamma\dot{\phi}$ into the equation of motion [3].

Multiplying through by $\dot{\phi}^2$ and taking the average over oscillations now gives a continuity equation,

$$\dot{\rho_\phi} + 3H\rho_\phi + \Gamma_\phi \rho_\phi = 0\,. \tag{2.50}$$

Energy is lost both by dilution from the expansion, but also the decay rate Γ_ϕ. The solution is

$$\rho_\phi = M^4 (a/a_{\rm osc})^{-3} \exp\left[-\Gamma_\phi (t - t_{\rm osc})\right]\,, \tag{2.51}$$

where 'osc' labels the time where coherent oscillations commence, and M^4 is the energy in the scalar field at this point. The field decays away as non-relativistic matter until the decay kicks in ($t \approx \Gamma_\phi^{-1}$), at which point it decays away exponentially.

The energy must go somewhere. Assuming it decays into relativistic matter,

$$\dot{\rho}_{\rm rad} + 4H\rho_{\rm rad} = \Gamma_\phi \rho_\phi\,, \tag{2.52}$$

where $\rho_{\rm rad}$ is the energy density of the produced radiation.

From the start of the coherent oscillations epoch, the universe behaves as if matter dominated, $a(t) \propto t^{2/3}$. Also, due the rapid expansion that has just taken place, there is no radiation. An approximate solution is then,

$$\rho_{\rm rad} \approx \frac{(6/\pi)^{1/2}}{10} m_{\rm pl} \Gamma_\phi M^2 \left(a/a_{\rm osc}\right)^{-3/2} \left[1 - \left(a/a_{\rm osc}\right)^{-5/2}\right]\,. \tag{2.53}$$

The energy density of relativistic decay products rapidly increases from 0 to about $m_{\rm pl}\Gamma_\phi M^2$ and thereafter decreases as $a^{-3/2}$, before eventually dominating over the ϕ field and decaying as a^{-4}, as radiation normally does.

2.5.6 The origin of density perturbations

While a full treatment is beyond the scope of this document, a rough sketch of this process is worthwhile. The production of the primordial density perturbations from inflation involves treating the field ϕ not just as a classical variable, but as a quantum operator (whose expectation value is roughly the classical variable we have been discussing so far). The fluctuations about this classical value, inherent to a quantum description, are what allow the production of density perturbations on cosmological scales.

In standard (inflationless) cosmology, any *comoving* length scale λ crosses the Horizon ($\sim H^{-1}$) only once. As the horizon gets steadily larger, any comoving scale initially within the horizon eventually becomes smaller than (crosses) the horizon at a later time (due to the growth of the horizon).

With inflation, this changes. All of the cosmologically important scales begin sub-horizon sized, cross outside the horizon during inflation, and then eventually cross back inside the horizon at later epochs. In some sense the entire purpose of inflation is to generate this sort of behaviour to solve the earlier problems, but a secondary effect is the impact it has upon quantum fluctuations in the scalar field.

In a rough sense, when these quantum fluctuations find themselves brought from sub-horizon to super-horizon scales, they can no longer be treated as quantum fluctuations any longer (as no behaviour can be correlated on scales past the horizon). Instead they 'freeze in' as classical perturbations to the metric, and re-enter the horizon as density perturbations.

Rather than delving too deeply into quantum field theory, we will simply take the result that the *mean square* fluctuation in ϕ is given by

$$\delta\phi \approx \frac{H}{2\pi}\,. \tag{2.54}$$

Naturally this causes a change in the energy density in ϕ, ultimately resulting in perturbations to the metric of spacetime (the analogous to the Newtonian gravitational potential).

The root mean square fluctuations σ about the gravitational potential Φ are related as,

$$\frac{d\sigma^2\Phi}{d\ln k} \equiv d_{\rm H}^2 = \frac{H^4}{\left(2\pi\dot{\phi}\right)^2}\,. \tag{2.55}$$

In exact de Sitter space inflation, H and $\dot{\phi}$ are both constant, which predicts are a perfectly scale invariant power spectrum ($\delta_{\rm H}$ = constant with scale). Fluctuations at the horizon size are the same at all times in de Sitter space; as different modes leave the horizon, they all leave with the same characteristic amplitude.

With observations from the CMB of the value of the earliest perturbations, we can constrain some inflationary parameters. Consider again the number of e-foldings,

$$N = \int H{\rm d}t = \int H{\rm d}\phi/\dot{\phi} = \int 3H^2{\rm d}\phi/V'\,. \tag{2.56}$$

Using an example $V = \lambda\phi^4$, then $N = H^2/(2\lambda\phi^2)$. This means (liberally using the slow-roll approximation where needed)

$$\delta_{\rm H} \sim H^2/\dot{\phi} = 3H^3/V' \sim \lambda^{1/2}N^{3/2}\,. \tag{2.57}$$

The observed $\delta_{\rm H} \sim 10^{-5}$, while we require 60 e-folds of inflation, putting a limit on this parameter,

$$\lambda \lesssim 10^{-15}\,. \tag{2.58}$$

For the quadratic potential, $V = m^2\phi^2$, so

$$m \lesssim 10^{-5} m_{\rm pl}\,. \tag{2.59}$$

On the other hand, inflation is not perfectly de Sitter (characterised by the slow-roll conditions). The predicted (and observed) spectrum of perturbations is not precisely scale-invariant, usually described through the *tilt*:

$$\text{tilt} \equiv 1 - n \equiv -\frac{d\ln\delta_{\rm H}^2}{d\ln k}\,. \tag{2.60}$$

Since the scale k crosses the horizon when $a/k = H^{-1}$, and H is nearly constant, derivatives with respect to k can be replaced by derivatives with a. The slow-roll can eliminate $\dot\phi$ for V', the Friedmann equation can eliminate H for V, and the horizon-scale amplitude $\delta_{\rm H}$ is known. All that remains is to replace V' and higher derivatives with slow-roll parameters, leading to:

$$1 - n = 6\epsilon - 2\eta\,. \tag{2.61}$$

When evaluated on the largest scales, 60 e-folds before the end of inflation, this gives predictions for $n = 0.97$ for quadratic potentials, and $n = 0.95$ for quartic potentials - very close to scale-invariance, but not quite.

Inflation has gone far beyond its original remit, encouraging research in this area. A huge number of models now exist, in widely varying degrees of complexity. When we turn to look at models of dark energy, we shall soon see how the framework of inflation has influenced the model building in this area also.

2.5.7 Topological defects

A very interesting, and much studied, aspect of scalar fields are topological defects. These are examples of inhomogeneities in the field brought about by a broken symmetry (either continuous or discrete).

The simplest to understand is the case of *domain walls*, which can be generated by a symmetry breaking potential,

$$V(\phi) = \frac{\lambda}{4}\left(\phi^2 - \eta^2\right)^2\,. \tag{2.62}$$

For $\phi^2 \gg \eta^2$, this is just a quartic potential. But for smaller values of ϕ, the potential actually has a double-well form, with a maximum at $\phi = 0$ and a minima either side.

The solution to the field equations with one spatial and one time direction is

$$\phi(x) = \eta\tanh\left(\sqrt{\lambda/2}\eta x\right)\,. \tag{2.63}$$

This is the so called 'ϕ^4-kink' solution, which centred on $x = 0$, takes ϕ from $-\eta$ at $x = -\infty$ to η at $x = \infty$. This solution is time-independent, and actually possess finite energy associated with the displacement of the field. So this solution describes an interpolation between regions of space with different values of ϕ (or different 'vacuums').

The Kibble Mechanism

These walls are time-independent - so they cannot decay, nor form in the first place, as currently described. However, should η be a function of time (for instance, by being related to the value of another, rolling scalar field, or the result of temperature dependent corrections to the potential), then the value of η can change from less than zero (creating a minimum at $\phi = 0$), to greater than zero (leaving $\phi = 0$ as an unstable maximum). The field ϕ will then begin falling into one or the other of these ground states, but the choice of minimum will depend on random fluctuations in ϕ, which can be expected to be different in separate regions of space. The 'kink' solution describes the transition of ϕ from one volume of space to a neighbouring region where ϕ has found itself in the alternate minimum. With added dimensions, this becomes a domain wall rather than just a one dimensional kink.

As they form, regions have a size (walls of a separation) of approximately the correlation length ζ. This can depend upon the particular conditions by which the symmetry is broken, but because if the causal horizon in cosmology, this sets a lower limit $\zeta < d_{\mathrm{H}}$. This was first pointed out by Kibble [85].

An effective 'fluid' of domain walls has $w = -2/3$, which is unfortunately too far from $w = -1$ to satisfy the observational constraints for dark energy.

More complicated topological defects (such as strings and textures) can be created from other sorts of symmetry breaking. See [145] for a further discussion of these and their cosmological implications.

Chapter 3

Models of Dark Energy

3.1 Problems with the standard model

Before giving a brief overview of the the plethora of dark energy models that exist in the literature, some discussion of what has motivated the sheer number of proposed alternatives to the standard model.

3.1.1 Fine tuning problem

In principle, there should be a contribution to the cosmological constant from the vacuum energy associated with the zero-point energy from quantum fields. The simplest way to understand this is that a quantised free (non-interacting) field is modelled as a collection of harmonic oscillators, one at each point in spacetime. Like any quantum mechanical harmonic oscillator, the lowest energy state is not zero, but some finite value above this.

To be more precise, this 'zero-point' energy of a quantum field with mass m is given by a sum over all the Fourier modes k,

$$\rho_{\rm vac} = \frac{1}{2}\int_0^\infty \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^2} \tag{3.1}$$

$$= \frac{1}{4\pi^2}\int_0^\infty \mathrm{d}k\, k^2\sqrt{k^2+m^2}\,. \tag{3.2}$$

This actually formally diverges: $\rho_{\rm vac} \propto k^4$. This understandable; there are an infinite number of harmonic oscillators if we assign one to every point in spacetime. However, no quantum field theory is expected to be valid up to infinite energy scales, and so an upper cut-off is usually added to the integral at the energy scale where our trust in the field theory breaks down. This leaves

$$\rho_{\rm vac} \approx \frac{k_{\rm max}^4}{16\pi^2}\,. \tag{3.3}$$

If trusted up to the Planck scale, $k_{\rm max} = m_{\rm pl}$, the energy density is huge:

$$\rho_{\rm vac} \approx 10^{74}\,{\rm GeV}^4\,. \tag{3.4}$$

This is about 10^{121} orders of magnitude larger than the observed value, an absurd result (though perhaps its absurdity is exaggerated by using the energy density, rather than the associated energy scale). Even using the energy scale of QCD for $k_{\rm max}$ results in a vacuum energy of $\rho_{\rm vac} \approx 10^{-3}{\rm GeV}^4$, still far too large.

Usually in quantum field theory, this background energy is ignored. It has no effect upon the predictions of the theory (where only the energy relative to some fixed background is important). It is only when trying to interpret this vacuum energy in the context of general relativity that the difficulties begin.

3.1.2 The coincidence problem

A number of other authors have noted a separate concern with the presence of a cosmological constant. Due to the density of matter decaying as R^{-3} while the vacuum energy density is unchanged by expansion, there is only a single period of time in the history of the universe when $\Omega_m \sim \Omega_w$. It is a curious fact that we find ourselves doing cosmology at just this epoch; earlier observers would have struggled to detect vacuum energy at all, while future observers will struggle to do cosmology at all, due to the exponentially accelerating expansion of the universe.

Different authors regard these facts with different levels of severity. Some consider it an unlikely coincidence that supports the need for a form of dynamical dark energy that neatly explains the apparent coincidence. Others are untroubled by it, mainly because they consider the coincidence to have no great significance (nor require explanation).

Finally, note before moving on, Weinberg argued that we should expect a cosmological constant of $\Omega_w \sim 1$ purely from anthropic grounds, long before the observational evidence was in support of this conclusion [150].

3.2 Models of dark energy

Regardless of motivation, it can be useful to have alternative models simply for the purposes of testing the standard model. We'll now look at a few of the most common types.

3.2.1 Phenomenological models of dark energy

The most simple approach is to treat dark energy as any other fluid with density $\rho_{\rm DE}$ and equation of state parameter $w_{\rm de}$. This approach does not even attempt to explain the nature of dark energy beyond a phenomenological background level. This leaves many features unexplained or undefined. For instance, when moving beyond a background level to examine perturbations, an important quantity is the sound speed $c_{\rm s}$. Without knowing the details of the fluid in a more fundamental fashion, this remains an entirely unknown parameter.

Despite all this, let us see how such fluids evolve with density.

Recalling the continuity equation,

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + P\right) = \dot{\rho} + 3\frac{\dot{a}}{a}\left(1 + w\right)\rho = 0\,, \tag{3.5}$$

we can write down a general solution for any w, even if it should be time (scale-factor) dependent:

$$\rho = \rho_0 \exp\left[-\int 3(1+w)\frac{{\rm d}a}{a}\right]\,. \tag{3.6}$$

For constant w, we have already seen this leads to power law solutions,

$$\rho \propto a^{-3(1+w)}\,. \tag{3.7}$$

We can also return to the acceleration equation,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(1 + 3w\right)\,. \tag{3.8}$$

If dark energy is to explain the accelerated expansion in this way, it needs to have $w < -1/3$ today, to drive $\ddot{a}/a$ positive.

At a background level, knowing the value of w and how it varies in time is sufficient to understand dark energy entirely. Unfortunately, observations are far from able to constrain an entire function, and so some simplification is required. An appropriate parametrisation for the equation of state parameter is a difficult question in itself.

3.2.2 Parametrisation of w

There have been many parametrisations proposed for the equation of state. Expansions of the form

$$w(a) = \sum_{n=0} w_n x_n(a)\,, \tag{3.9}$$

are very common, here the expansion functions are many and varied in the literature.

We have already considered constant w: $x_0(a) = 1; x_n = 0, n \geq 1$. Some other alternatives include power laws in redshift z, so that $x_n(z) = z^n$, or in scale factor with $x_n(z) = (1 - R/R_0)^n = (z/(1+z))^n$. A less common but possible parametrisation is to use a logarithmic approach, $x_n(z) = [\log(1+z)]^n$.

These series are all usually truncated at the second term, as more than two free parameters becomes difficult to constrain with the available data. Fortunately, for many models over a modest range in redshift, two parameters is quite often sufficient to capture the dynamics.

A very typical two-parameter model is from the second case, with $w = w_0 + w_1 z$. This Taylor expansion was introduced by Huterer and Turner [77], as well as Weller and Albrecht [151].

This fit does have an flaw, in that at high enough redshift, for $w_a \neq 0$, the second term will obviously dominate and become very large. In fact, taken at face value this parametrisation is excluded by constraints of of BBN [79]. This makes it difficult to combine low-redshift observations (for instance, SN 1a data) with high redshift (such as the CMB).

A third case with $w(z) = w_0 + w_1 z/(1+z)$ was introduced by Chevallier and Polarski [35] and Linder [99]. This fit (not a Taylor expansion, as Linder emphatically points out) is a simple enough parametrisation to be constrained by the data, but has been shown to successfully fit a wide variety of scalar field dark energy models (that we shall shortly be considering). The additional $(1+z)$ factor in the denominator of the second term means that it is well behaved at high redshift.

3.2.3 Scalar field models of dark energy

Casting around for the physical origins of dark energy, many authors have drawn inspiration from inflation, and attempted to construct scalar field models that generate the desired behaviour. We have already seen in the case of inflation, with an appropriate potential $V(\phi)$ and initial conditions, the field can be made to behave as a fluid with an equation of state close to $w = -1$. In particular, the dynamical nature of the field

brings hope that the coincidence and fine-tuning problems might somehow be either eliminated or alleviated.

These scalar field models have not (as yet) met with the same success as inflation, primarily because of the limited data and lack of unique observational signatures. Combined with the great interest in solving the puzzle of the dark energy, this has lead to a huge variety of scalar field models.

Quintessence

Perhaps the most basic scalar field dark energy model and one of the earliest, is dubbed *quintessence* [114, 152, 30]. This is described by a canonical scalar field ϕ with a potential $V(\phi)$ minimally coupled to gravity, with the goal being that it naturally leads to late time inflation. Just as with inflation, the action is,

$$S = \int \mathrm{d}^4\sqrt{-g}\left[-\frac{1}{2}\left(\nabla\phi\right)^2 - V(\phi)\right], \tag{3.10}$$

with $\left(\nabla\phi\right)^2 = \partial_\mu\phi\partial^\mu\phi$. In a universe dominated by the scalar field (roughly the case today), the Friedmann equation gives

$$H^2 = \frac{8\pi G}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right] \tag{3.11}$$

and the acceleration equation gives

$$\frac{\ddot{a}}{a} = -\frac{8\pi G}{3}\left[\dot{\phi}^2 - V(\phi)\right]. \tag{3.12}$$

The universe accelerates provided $\dot{\phi}^2 < V(\phi)$. Like inflation, a flat potential is required.

Unlike inflation, the situation is complicated by the non-negligible dark matter present today, and by its dominating impact upon the earlier evolution. The slow-roll parameters used for inflation are therefore no longer a good indicator of an inflationary solution existing, and a redefinition in terms of H such as $\epsilon = -\dot{H}/H^2$ is more suitable to check for an accelerated expansion.

Recall just as with inflation, the equation of state for ϕ is given by

$$w_\phi = \frac{P}{\rho} = \frac{\dot{\phi}^2 - 2V(\phi)}{\dot{\phi}^2 + 2V(\phi)}, \tag{3.13}$$

limiting the equation of state with $-1 \leq w_\phi \leq 1$.

It can be interesting to work out the potential needed to produce produce an expansion described by a power law when the field dominates the universe,

$$a(t) \propto t^p. \tag{3.14}$$

For $p > 1$, there is accelerated expansion. The potential ϕ needs to be rolling down to produce this power law behaviour is an exponential,

$$V(\phi) = V_0 \exp\left(-\sqrt{\frac{16\pi}{p}}\frac{\phi}{m_{\rm pl}}\right) , \tag{3.15}$$

with V_0 as a constant to keep the dimensions correct. The field evolves with $\phi \propto \ln t$.

Beyond the fact that an accelerated expansion can be produced, these potentials also allow for cosmological *scaling solutions*, the term given to solutions where the dark energy density is proportional to the dark matter energy density.

Since $p = 1$ in the exponential potential is the boundary between accelerating and decelerating expansions, this suggests potentials less steep than this exponential potential might to produce acceleration. Indeed, the original quintessence model used a power law potential,

$$V(\phi) = \frac{M^{4+\alpha}}{\phi^\alpha} , \tag{3.16}$$

with α as a (usually positive) constant, and M also a constant.

What was once the fine-tuning of the cosmological constant now resides in the choice of M for a given α. The qualitative behaviour of the universe is to enter a so called 'tracking' regime, where the energy density of the field catches up to the background matter (the state of the universe today) when $\phi \approx m_{\rm pl}$ [114, 30]. With $\rho_\phi \approx V(\phi)$, this gives

$$M = \left(\rho_\phi^{(0)} m_{\rm pl}^\alpha\right)^{1/(4+\alpha)} , \tag{3.17}$$

constraining the combination of α and M by matching to the observed dark energy density (superscript 0 as the convention for the value today). For example, with $\alpha = 4$, $M \sim 10^6$GeV. This energy scale is much closer to particle physics scales, but the mass scales involve are still incredibly tiny, and the field values uncomfortably large. There are no widely accepted particle physics models that include potentials of this form, and perhaps worse still, these potentials are generally non-renormalisable, making any approach to quantize the theory mired in divergences that cannot be dealt with in the usual way in quantum field theory.

From a more practical standpoint, a frustrating aspect to quintessence is the lack of any real firm predictions. With no real theoretical constraints to the potential, any particle dynamical $w(t)$ could in principle be matched by a 'designer' potential. This is a stark difference to inflation, where immediate success was found in the form of predicting the spectrum of primordial fluctuations. This problem is likely to persist until a much more fundamental theory for ϕ and its potential is discovered, or until the dark energy is found to match so well observationally to a cosmological constant.

K-essence

Quintessence relies on an arbitrary potential $V(\phi)$ in the action. Unsatisfied with merely this freedom, a more exotic approach to dark energy is to use a non-canonical kinetic part to the action, with

$$S = \int \mathrm{d}^4x\sqrt{-g}p(\phi, X)\,, \tag{3.18}$$

where $X \equiv -(1/2)(\nabla\phi)^2)$. Usually the form of the Lagrangian density is restricted to be of the following form [37, 10, 9]

$$p(\phi, X) = f(\phi)\hat{p}(X)\,. \tag{3.19}$$

Phantom fields

So far we have always assumed $w \geq -1$. However the observational data does not enforce this restriction, it is purely a theoretical prejudice. One example of a model that has $w \leq -1$ is a 'phantom' or ghost field, which replaces the canonical kinetic term in the action with the same term of negative sign,

$$S = \int \mathrm{d}^4\sqrt{-g}\left[\frac{1}{2}(\nabla\phi)^2 - V(\phi)\right]\,. \tag{3.20}$$

With the sign of the kinetic term reversed, the energy density and pressure become $\rho = -\dot{\phi}^2/2 + V(\phi)$ and $p = -\dot{\phi}^2/2 - V(\phi)$, so the equation of state becomes

$$w_\phi = \frac{p}{\rho} = \frac{\dot{\phi}^2 + 2V(\phi)}{\dot{\phi}^2 - 2V(\phi)}\,. \tag{3.21}$$

Therefore $w_\phi < -1$ for $\dot{\phi}^2/2 < V(\phi)$. For this case, the energy density actually *grows* as the universe expands. A universe dominated by a scalar field like this grows so fast that the Hubble rate and scalar curvature diverges, reaching infinite values within a finite time (a 'Big Rip' singularity), though this can be avoided if the potential has a maximum for the field to settle [134].

While possessing interesting properties for cosmology, phantom fields suffer from ultra-violet instabilities when the attempt is made to interpret them at a quantum level. It remains to be seen if these problems are insurmountable.

3.2.4 Modified gravity models of dark energy

$f(R)$ gravity

The alternative to modifying the matter side of the Einstein equations is to modify the gravitational side. This is usually done at a fundamental level by considering the

action for gravity [32, 33],

$$S = \int \mathrm{d}^4x\sqrt{-g}f(R)\,, \tag{3.22}$$

where $f(R)$ is an arbitrary function in terms of R, the Ricci scalar.

For general relativity, $f(R) = R/16\pi G$, but additional terms are possible.

Attempting to modify gravity in this fashion produces models (quite sensibly) known as $f(R)$ gravity models. The field equations an arbitrary function $f(R)$ produces are quite complicated, but can be simplified significantly by making a conformal transformation of the metric,

$$g^{(E)}_{\mu\nu} = e^{2\omega}g_{\mu\nu}\,, \tag{3.23}$$

where w is a smooth, positive function of spacetime coordinates. With an appropriate choice of form for ω, in this new metric (the Einstein frame, indicated by the superscript) then the action can be re-written as the general relativity action, plus an additional scalar field degree of freedom $\phi \propto \omega$, which has a canonical kinetic term and a potential that is a function of ϕ and R.

Rewriting gravity in this fashion is just a formal change, but does highlight the difficulty in distinguishing modifications to gravity from an unknown scalar field comprising the dark energy. The major distinguishing factor is not how the universe evolves, but instead how structure grows in the universe. Modifications to gravity typically leave changes in the perturbed Einstein equations that are difficult to mimic with simple fluid models (though see Chapter 8 where we shall see that modified gravity can be hard to distinguish from dark energy exchanging energy-momentum with the dark matter).

Modified Gravity from higher dimensions

Though we will not dwell on this subject, another modification to gravity involves including additional dimensions of spacetime. Two of the most well known ones are the Randall-Sundrum (RS) braneworld model [123] and the braneworld model of Dvali-Gabadadze-Porrati (DGP) [55]. In these scenarios, the four dimensional universe we inhabit (the *brane*) is embedded in a higher dimensional *bulk* spacetime, and the behaviour of gravity in the brane is impacted by how it leaks out into the bulk. These sorts of models can produce accelerated expansions, and provide an alternative to simply modifying the action in four dimensions.

Solar system constraints

A major factor in using modified gravity to explain the accelerated expansion is that solar system tests of gravity must also be accommodated. Changing gravity only on the largest of scales is not always easy, and almost inevitably leads to small scale changes in the behaviour of gravity that can be tested on a solar system level. Many modified gravity schemes fail to meet this constraint (see e.g. [36] for more details).

3.3 Further reading

The above is little more than a summary of the huge number of dark energy models that have been created. The interested reader is directed to a [46] for a substantial review on dark energy models.

Chapter 4

Hybrid Quintessential Inflation

4.1 Introduction

Peebles and Vilenkin proposed in [115] that both inflation and dark energy could be a result of the same scalar field, with vacuum expectation value ϕ, interacting only with gravity and itself via the potential term $V(\phi)$, which they chose to be

$$\begin{aligned} V(\phi) = &\quad \lambda\left(\phi^4 + M^4\right) \qquad \text{for} \quad \phi < 0 \\ = &\quad \frac{\lambda M^8}{\phi^4 + M^4} \qquad \text{for} \quad \phi \geq 0. \end{aligned} \tag{4.1}$$

At tree level the evolution of the vacuum expectation value of the ϕ-field is governed by

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = -\frac{\partial V}{\partial \phi}\,, \tag{4.2}$$

with the cosmological expansion rate related to energy density ρ by the Friedmann equation,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3m_{\rm pl}^2}\rho\,, \tag{4.3}$$

where the Planck mass $m_{\rm pl} = G^{-1/2} = 1.22 \times 10^{19}$ GeV and overdots signify derivatives with respect to coordinate time.

In the original scenario [115], the universe begins dominated by the potential energy of the scalar field ϕ. The field has some large, negative value, and slowly rolls toward

the origin. The slow-roll conditions of inflation are satisfied, and the universe expands exponentially. During inflation, fluctuations in ϕ are frozen into the field. These seed future structure in the universe. To provide the correct level of fluctuations, as usual $\lambda \simeq 10^{-14}$. When $\phi \sim -m_{\rm pl}$, the inflationary epoch draws to a close. Thus far the situation is identical to ϕ^4 chaotic inflation [96], but from here on it differs. The kinetic energy $\frac{1}{2}\dot{\phi}^2$ of the field is no longer negligible, and it soon begins to dominate the universe [136]. This phase is termed "kination". The field behaves approximately as stiff matter, with its energy density redshifting as $\rho_\phi \propto a^{-6}$. Taking the potential energy at the end of inflation to be the same as the kinetic energy, a simple estimate gives

$$H^2 \simeq \frac{8\pi}{3}\lambda m_{\rm pl}^2 \left(a_x/a\right)^6 \,, \tag{4.4}$$

with solution, $a \propto t^{1/3}$ and thus $t = (3H)^{-1}$. Here the subscript x indicates the value at the end of inflation. During this kinetic dominated phase, the field moves as

$$\phi = \sqrt{\frac{6}{8\pi}} m_{\rm pl} \ln\left(a/a_x\right) - m_{\rm pl} \,. \tag{4.5}$$

Thus, spacetime started in approximately de Sitter form, and ended dominated by the kinetic energy of a homogeneous field. Ford showed [64] this transition leads to gravitational particle production (see also [113] for a discussion of the *Unruh* effect, to which this is closely related). This effect is subtle, hinging upon the difficulty of defining the vacuum (quantum fields in their groundstates) in curved spacetime. In fact, the vacuum appears to depend upon the observer's path through spacetime. When the universe changes from de Sitter domination to being dominated by the scalar field (or an observer changes from uniform motion to accelerated), the vacuum requires redefinition, suddenly introducing a thermal background. In some sense the produced particles have been present all along; the change of spacetime geometry simply revealed their presence.

The result is a small but important energy density of relativistic particles,

$$\rho_{\rm r} \simeq 0.01 N_{\rm s} H_x^4 \,, \tag{4.6}$$

where $N_{\rm s}$ is the number of scalar fields. Thermalization was found in [115] to occur at a radiation temperature

$$T_{\rm th} \simeq 10^9 N_{\rm s}^{3/4} \text{ GeV} \,. \tag{4.7}$$

The radiation redshifts away slower than the energy in the field. Provided M in Eq. (4.1) is not too large, Peebles and Vilenkin showed the universe transitions to a radiation

dominated epoch, with a temperature of

$$T_{\rm RH} \simeq 10^3 N_{\rm s}^{3/4}\ {\rm GeV}\,. \tag{4.8}$$

The field remains essentially static from this point on, mimicking a cosmological constant. The value of M can be then chosen so that $V(\phi_{\rm r})$ matches today's observed value of dark energy density. In the analysis in [115], this turned out to be $M \sim 10^6$ GeV.

Unlike most inflationary scenarios, in the one by Peebles and Vilenkin [115] at the end of inflation the inflaton field does not undergo a series of damped oscillations about a potential minima, which is the mechanism by which reheating usually takes place [3, 52]. The absence of this behaviour means gravitational particle production has to be relied upon instead. The same gravitational mechanism works to produce a stochastic background of gravitational waves (GW) [137, 4, 126], and the overproduction of gravity waves is one of the potential dangers of this model. Gravitons behave as massless scalar fields, and the energy density of the gravitons by the end of inflation is just that of a single scalar field, times two polarization states, so that the ratio of the energy densities in GW to radiation at the end of inflation is simply $(\rho_{GW}/\rho_{\rm r})_x \simeq 2/N_{\rm s}$. At the time of nucleosynthesis ρ_{GW} will contribute as an effective extra degree of freedom in radiation, but the success of Big Bang Nucleosynthesis (BBN) in predicting the abundances of light elements puts a constraint on the extra number of light degrees of freedom, be those neutrinos species or gravitons, such that $(\rho_{GW}/\rho_{\rm r})_{BBN} < 0.2$; tracing back the evolution of this ratio up to the start of the kination period this would set a lower bound $N_{\rm s} \gtrsim 100$ [115].

The problem of gravity wave overproduction is quite generic in models of inflation followed by a long period of kination [142], or in general stiff matter domination [67], like brane world inflation [104, 45, 76]. To allow for a more effective reheating process, able to suppress the relative contribution of the GW at the time of BBN, one can invoke alternative methods like instant preheating [31, 128], curvaton reheating [60, 50, 95], or Born-Infeld reheating [127]. Any alternative implies introducing extra scalar degrees of freedom at the time of inflation, like in curvaton reheating, and/or direct couplings of the inflaton field to the light degrees of freedom as in instant preheating.

In this chapter we explore the possibility of a simpler scenario, recovering a more typical reheating mechanism driven by the decay of an oscillating massive field [3, 52]. We extend Peebles and Vilenkin model by a new scalar field χ coupled to the inflaton field with a hybrid-like potential [98]. In our scenario, once the inflaton field falls below

a critical value, the χ field can start oscillating, thus gaining energy that afterwards can be converted into radiation through perturbative decay, as in the usual reheating mechanism. However, in contrast to the standard reheating picture, in our scenario reheating takes place during kination instead of the more standard matter domination. Unless the perturbative decay of the χ field is tiny, this results in a larger reheating T than in the original model of Peebles and Vilenkin [115], and a shorter kination phase. The more efficient reheating also ensures that radiation domination takes over kination, well before the inflaton vacuum energy starts dominating again.

This chapter is organized as follows. In Section II the potential and parameters of the model are set. Also the general behaviour is described of the field χ after inflation, when it can oscillate and drive reheating. The reheating temperature $T_{\rm RH}$ is then computed in Section III. For the mechanism to work, we need to check first that χ indeed oscillates when the inflaton field passes through the critical point, and that it does not backreact on the evolution of ϕ. Fulfilling these conditions sets the constraints on the model parameters, which are given in Section IV. In Section V we present the range of $T_{\rm RH}$ consistent with the constraints. The transition to a radiation dominated universe is studied in Section VI. Once the constraints are fulfilled, the transition before the onset of dark energy domination is practically ensured. As this is a hybrid-like model, we comment on the issue of domain walls in section VII. Finally in Section VIII we present the summary and future work related to quantum corrections.

4.2 General behaviour of the hybrid field

To enable once more the standard reheating, we introduce a new scalar field χ, that we couple to the inflaton field. The effective potential at tree-level is

$$U(\phi,\chi) = V(\phi) + \frac{g^2}{2}\chi^2\left(\phi^2 - m^2\right) + \frac{\lambda_\chi}{4}\chi^4\,, \tag{4.9}$$

where $V(\phi)$ is the potential given by equation (4.1). The parameters g, m, and λ_χ are as yet undetermined constants. Note that when χ is relaxed near the origin, $U(\phi,\chi) \approx V(\phi)$. A plot of the potential is shown in Figure 4.1.

The χ-field is assumed to be located somewhere near its minima at $\chi = 0$. As the ϕ field evolves from large negative values to large positive ones, the turning point at the origin temporarily becomes unstable for the χ-field, and two minima are generated on either side. The position of the minima, $\chi_{\rm min}$ are given by

$$\chi^2_{\rm min} = \frac{g^2}{\lambda_\chi}\left(m^2 - \phi^2\right)\,. \tag{4.10}$$

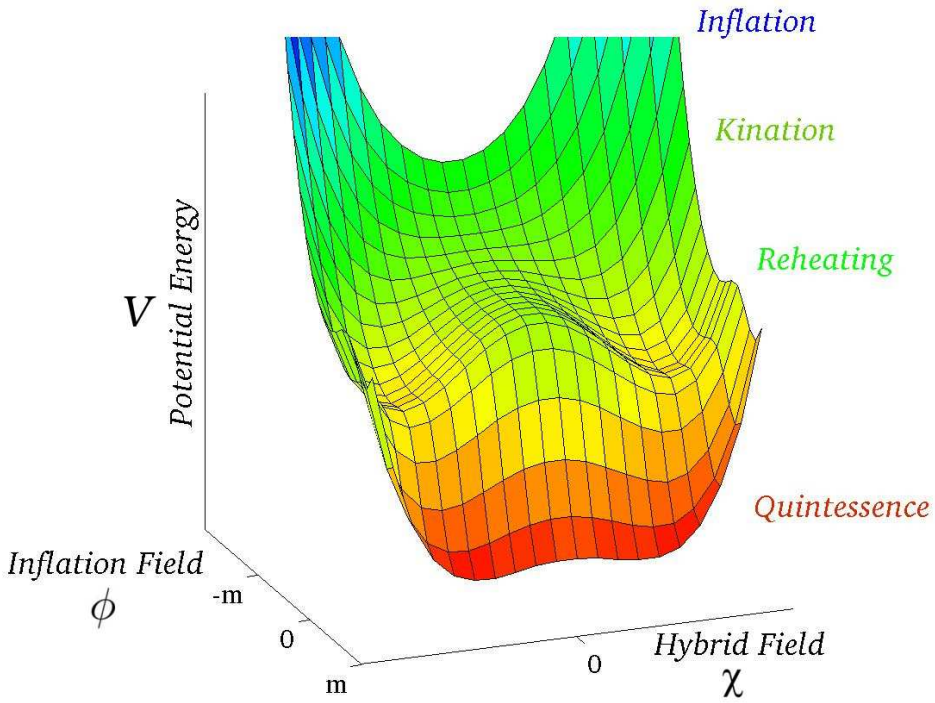

Figure 4.1: A plot of the potential, with the different cosmological regimes labelled along the ϕ axis. The 'hybrid' field χ gains kinetic energy by falling into the wells of the potential between $\phi = -m$ and $\phi = +m$.

They exist only while $|\phi| < m$. The bottom of the well is at a negative value of potential energy. This is a result of how the energy of the system has been defined. It has no physical consequence for the present scenario, since the total energy density will always remain positive, as it is dominated by the kinetic inflaton energy density.

The χ-field has a characteristic response time τ to react to changes in the potential. This can be estimated as $\tau^{-1} \sim m_\chi(\phi) \sim g(m^2 - \phi^2)^{1/2}$, provided ϕ is not too close to m. If changes in the potential take place quicker than this response time, the field will have no time to react. For instance, if ϕ moves between $-m$ and $+m$ in a time $\Delta t \ll \tau$ throughout, χ will have no time to move before the origin becomes a stable minima once more. On the other hand, if the changes take place on timescales that are much longer than the response time, i.e. $\Delta t \gg \tau$, the χ field will be able to relax into the minima very quickly. If this is the case throughout, there will hardly be any oscillations (and hardly any reheating).

A fact alleviates the above difficulties: ϕ is slowing down, and is doing so quickly (its evolution is logarithmic in time). This means the timescale Δt becomes gradually longer. Furthermore, the timescale τ that governs how quickly χ reacts is not static.

It is longest when $|\phi| \approx m$ and is shortest at $\phi = 0$. We will take the view that the field begins to move before $\phi \approx 0$. This places an immediate constraint on the model parameters: $\Delta t \gg \tau|_{\phi=0}$. We should immediately note the better this inequality is satisfied, the sooner χ will start to move after becoming unstable. This will limit the amount of energy for its oscillations.

As a simplified picture, consider the χ field to be frozen at the origin up until a time τ after becoming unstable. Afterwards, the potential will be approximately constant during the fast oscillations of the field. The energy in the oscillations (available for reheating) will then be the depth of the potential well at the point when the field begins to move. This will be some fraction $f(g, m)$ of the maximum potential well depth. The energy available for reheating can then be written:

$$\rho_\chi^{(0)} \simeq f\,(g, m)\,\frac{m^4 g^4}{4\lambda_\chi}\,. \tag{4.11}$$

Eventually, after many oscillations, ϕ reaches $+m$, and the χ - field then oscillates about its stable minima at the origin. These oscillations will reheat the universe.

The fraction $f(g, m)$ now needs to be estimated. For this, we assume the χ -field moves some short time after it becomes unstable, with the instability occurring at $\phi = -m$. As this time signals the start of reheating, it will be called $t_{\rm re}$. Writing the time since the start of reheating as δt, then we can estimate that the χ field will begin to move when $\tau \sim \delta t$. In other words, the field moves after it has been unstable for a time approximately the same as its characteristic reaction time (which is itself a function of δt). The depth of the well is given by

$$U_{\rm min} = m^{-4}\left(m^2 - \phi^2\right)^2 \frac{m^4 g^4}{4\lambda_\chi}\,. \tag{4.12}$$

Comparison with equation (4.11) shows

$$f(g, m) = m^{-4}\left(m^2 - \phi^2\right)^2\,, \tag{4.13}$$

with ϕ evaluated when $\delta t \simeq \tau$. We assume ϕ moves only slightly, an amount $\delta\phi = \phi - m$, before χ begins to move. Re-writing equation (4.5) gives

$$\delta\phi \approx \frac{1}{3}\sqrt{\frac{6}{8\pi}}\, m_{\rm pl}\frac{\delta t}{t_{\rm re}}\,. \tag{4.14}$$

We wish to find the value of δt satisfying $\delta t \simeq \tau$. As $\tau^{-1} \simeq g(m^2 - \phi^2)^{1/2}$, we can insert $\phi = \delta\phi - m$ to find

$$\tau^{-2} \sim m g^2 \frac{2}{3}\sqrt{\frac{6}{8\pi}}\, m_{\rm pl}\frac{\delta t}{t_{\rm re}}\,. \tag{4.15}$$

The same substitution into $f(g,m)$ yields

$$f(g,m) \simeq 4\frac{(\delta\phi)^2}{m^2}\,, \tag{4.16}$$

to leading order in $\delta\phi$.

Applying the condition $\delta t \simeq \tau$, we can find the value of δt when the χ - field begins to move. Relating $\delta\phi$ to δt with equation (4.14), we finally obtain

$$f(g,m) \simeq \left(\frac{1}{3\pi}\right)^{2/3} \left(\frac{m_{\rm pl}}{m}\right)^{8/3} (m_{\rm pl} t_{\rm re})^{-4/3}\, g^{-4/3}\,. \tag{4.17}$$

Note that this order of magnitude approximation breaks down if χ begins moving when $\delta\phi$ is not small compared to m, signalled by $f(g,m)$ approaching (or exceeding) unity. This expression should not be trusted in such circumstances. Assuming $\delta\phi$ to be small is a slightly stronger constraint than the constraint discussed earlier, that $\Delta t \gg \tau$. The former gives the condition the field moves *quickly* after becoming unstable, the latter is just the condition the field moves *at all*, at or before $\phi \approx 0$. We will return to this point when discussing the parameter constraints in Section IV.

4.3 Reheating temperature

Once $\phi > m$, the χ field will return to the origin. If it acquired kinetic energy due to its temporary displacement, it will now oscillate about the origin. The previous section established the amount of energy expected from these oscillations.

If the oscillations are small, the field can be approximated as undergoing simple harmonic motion. In such a case, the elementary theory of reheating can be applied [3, 52]. In this phenomenological approach, an extra term $\Gamma_\chi \dot{\chi}$ is added to the equation of motion of the field to account for particle decay. The value of Γ_χ is taken to be the decay rate of the particle. The field χ then obeys the equation of motion

$$\ddot{\chi} + 3\frac{\dot{a}}{a}\dot{\chi} + \Gamma_\chi \dot{\chi} = -\frac{\partial U}{\partial \chi}\,. \tag{4.18}$$

This approach is valid when the oscillations are small and the oscillations are well approximated as simple harmonic motion. For a more complete picture, valid at the early stages we should also consider the effect of preheating [86, 87] (see also [31, 128] for an analysis of preheating in the context of quintessential inflation). However, these details will be ignored in this paper and we will examine only the simplest reheating estimates.

We stress that this phenomenological approach is only valid while the field is undergoing coherent oscillations about its minima. It will not be valid otherwise, nor over short timescales, and is likely to fail when $|\phi| \sim m$. For this reason we will consider reheating to take place only while $\phi > m$, so that we may have confidence our calculations are always carried out in an appropriate regime.

We assume H, ϕ and $\dot{\phi}$ can be taken to be approximately constant over a single oscillation. Equation (4.18) is re-written by replacing $\dot{\chi}^2$ by its value over a complete oscillation, $\langle\dot{\chi}^2\rangle_{cycle} = \rho_\chi$, which is valid for simple harmonic motion. This yields

$$\dot{\rho_\chi} + 3H\rho_\chi + \Gamma_\chi\rho_\chi = g^2\langle\chi^2\rangle\phi\dot{\phi}\,. \tag{4.19}$$

Assuming the field is still undergoing simple harmonic motion, $\frac{1}{2}\rho_\chi = V(\phi,\chi) = \frac{1}{2}g^2\phi^2\langle\chi^2\rangle$. Then we can write

$$\dot{\rho_\chi} + 3H\rho_\chi + \Gamma_\chi\rho_\chi = \frac{\dot{\phi}}{\phi}\rho_\chi\,. \tag{4.20}$$

This can be solved,

$$\rho_\chi = \rho_\chi^{(m)}\left(\frac{a_m}{a}\right)^3\frac{\phi(t)}{m}\exp\left[-\Gamma_\chi(t-t_m)\right], \tag{4.21}$$

with

$$\frac{\phi}{m} = \sqrt{\frac{6}{8\pi}}\frac{m_{\rm pl}}{m}\ln\left(\frac{a}{a_m}\right) + 1\,. \tag{4.22}$$

We have used subscript m to indicate the value of a variable when $\phi = m$. From energy conservation, it follows that the radiation density must obey

$$\dot{\rho}_r + 4H\rho_{\rm r} = \Gamma_\chi\rho_\chi\,. \tag{4.23}$$

Imposing the condition there is no radiation at the start of decay, an approximate solution is found by neglecting the exponential decay of ρ_χ. This will be valid up to $t \approx \Gamma_\chi^{-1}$. After this, the energy in the χ field will decay rapidly away and the radiation will simply redshift with its usual a^{-4} behaviour.

Inserting our earlier expression for ρ_χ into equation (4.23), the solution for the radiation energy density can be written as

$$\rho_{\rm r} = \frac{3}{4}\rho_\chi^{(m)}\Gamma_\chi t_m\Big[\quad\big(1-(t/t_m)^{-4/3}\big)\big(1-b\big) + b\,(4/3)\ln(t/t_m)\Big], \tag{4.24}$$

with $b = \frac{1}{4}\sqrt{\frac{6}{8\pi}}\frac{m_{\rm pl}}{m}$ and the constant of integration chosen so that there is no radiation at $\phi = m$.

Assuming $(t/t_{\rm re})^{-4/3} \ll 1$ before t reaches Γ_χ^{-1}, the energy density in radiation should be well approximated by

$$\rho_{\rm r} \simeq \frac{3}{4}\rho_\chi^{(m)}\Gamma_\chi t_m \left[1 - b + \frac{4}{3}b\ln\left(t/t_m\right)\right] . \tag{4.25}$$

The energy density of the radiation continues growing logarithmically, despite the energy loss from redshifting. This is due to the mild amount of energy being added to the χ field by its coupling to ϕ. It will continue to grow in this way until $t \approx \Gamma_\chi^{-1}$. If this is a sufficiently late time that the last term in Eq. (4.25) dominates, then

$$\rho_{\rm r} \sim \frac{1}{4}\sqrt{\frac{6}{8\pi}}\frac{m_{\rm pl}}{m}\rho_\chi^{(m)}\Gamma_\chi t_m \ln(t/t_m) . \tag{4.26}$$

Once the universe has had time to thermalize, the temperature is related to the energy density via,

$$\rho_{\rm r} = \frac{g_*\pi^2 T^4}{30} , \tag{4.27}$$

where g_* is the number of degrees of freedom.

Thermalization occurs when the interaction rate $n_r\sigma$ becomes comparable to the expansion rate H, where n_r is the number density of the light degrees of freedom and σ their interaction cross section. We can estimate that the light degrees of freedom are created with a typical energy $\omega \sim \rho_{\rm r}^{1/4}(a_{\rm re}/a)$, and $\sigma \sim \alpha_g/\omega^2$, with α_g being the strength of the mediating interactions. Using Eq. (4.26) with the expression of $t_{\rm re} \simeq t_m$ given in the next section Eq. (4.31), and as a typical value for the coupling in the cross-section $\alpha_g \sim 0.01$, one gets that at the beginning of the reheating period,

$$\frac{n_r\sigma}{H_m} \sim \frac{\rho_{\rm r}^{1/4}\alpha_g}{H_m} \sim 10^5(10\alpha^{1/4})\left(\frac{m}{m_{\rm pl}}\right)^{1/3}\left(\frac{g^{11/3}}{\lambda_\chi}\right)^{1/4} . \tag{4.28}$$

We have also written the decay rate for a massive particle as $\Gamma_\chi \simeq \alpha m_\chi$, where α is the coupling constant mediating the decay, and $m_\chi \simeq gm$ the χ mass. For the analyses of the reheating done in this section, the decay rate and therefore the coupling α must be such that $\Gamma_\chi t_m \ll 1$. For values of the parameters consistent with the constraints given in the following section, this is general the case with $\alpha \simeq 10^{-4}$. Significantly, with this choice for α then Eq. (4.28) is already larger than one. Therefore, light degrees of freedom thermalized promptly after they are produced.

We have confirmed these approximations are successful in their appropriate regimes by numerically solving this system of differential equations (Friedmann's equation, the equations of motion for both fields, and the radiation energy density). Figure 4.2 shows the numerically determined evolution of the χ - field as it becomes unstable, and the subsequent damped oscillations that reheat the universe.

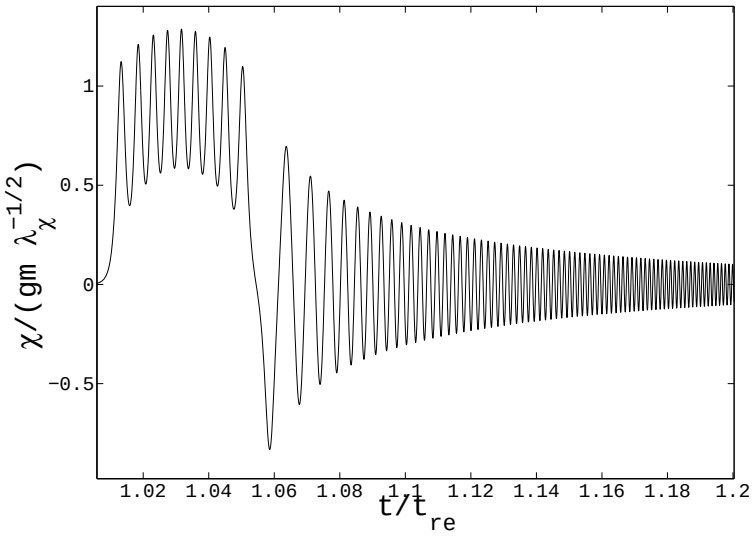

Figure 4.2: Numerical solution for the motion of the χ field, with $g = 5 \times 10^{-4}$, $\lambda_\chi = 1$, $m = 5 \times 10^{16}$ GeV and $\alpha = 1 \times 10^{-4}$. The field begins at rest, with a small displacement from the origin. The resulting oscillations continue when the field returns to oscillate about the origin, and are well approximated as damped, simple harmonic motion.

4.4 Parameter constraints

We have made two assumptions that can be formulated as simple constraints on combinations of parameters. The first of these is that χ begins to move toward its new equilibrium at or before $\phi \approx 0$. Earlier we noted this condition was $\Delta t \gg \tau$. The second assumption is that the χ field does not significantly influence the motion of the ϕ field. It could do this either from the coupling, or from the energy density of the field modifying the expansion rate of the universe.

First we shall calculate Δt. As $\phi = -m$ at $t_{\rm re}$, we can use equation (4.22) with these values inserted. Re-arranging,

$$t_m = t_{\rm re} e^{\sqrt{48\pi}\frac{m}{m_{\rm pl}}}\,, \tag{4.29}$$

and by writing $\Delta t = t_m - t_{\rm re}$, gives

$$\Delta t = t_{\rm re}\left(e^{\sqrt{48\pi}\frac{m}{m_{\rm pl}}} - 1\right) \simeq t_{\rm re}\sqrt{48\pi}\frac{m}{m_{\rm pl}}\,, \tag{4.30}$$

with the rightmost expression in the case $m/m_{\rm pl} \ll 1$. Thus, it is noteworthy that $t_{\rm re} \approx t_m$ is a good approximation.

An estimate of $t_{\rm re}$ is still needed. We use equation (4.4) and $H = \frac{1}{3}t^{-1}$, to write it

in terms of a ratio of scale-factors,

$$t_{\rm re} \simeq \left(\frac{a_{\rm re}}{a_x}\right)^3 \frac{1}{\sqrt{24\pi\lambda m_{\rm pl}^2}}\,, \tag{4.31}$$

where $a_{\rm re}$ is the scale factor at $t_{\rm re}$. The time for ϕ to move between $-m$ and $+m$ is then given by

$$\Delta t \simeq \left(\frac{a_{\rm re}}{a_x}\right)^3 \sqrt{\frac{2}{\lambda}}\frac{m}{m_{\rm pl}^2}\,. \tag{4.32}$$

The ratio of scale-factors can be found from equation (4.5). Re-arranging and setting $\phi = -m$ at $a = a_{\rm re}$,

$$a_{\rm re}/a_x = e^{\left(1-m/m_{\rm pl}\right)\sqrt{\frac{8\pi}{6}}} \simeq 8\,. \tag{4.33}$$

As such the time interval available can be written simply as

$$\Delta t \;\simeq\; 10^{-9}\frac{m}{m_{\rm pl}}\,(\text{GeV})^{-1}\,, \tag{4.34}$$

with $\lambda = 10^{-14}$. Recalling $\Delta t \gg \tau$, a constraint on the model parameters can be constructed:

$$\left(\frac{m}{m_{\rm pl}}\right)^2 g \gg\; 10^{-10}\,. \tag{4.35}$$

Now that we have an expression for $t_{\rm re}$, a constraint on $f(g,m) \ll 1$ can also be calculated. Replacing Eq. (4.31) into (4.17) gives

$$\left(\frac{m}{m_{\rm pl}}\right)^2 g \gg 2\sqrt{2\lambda}\left(\frac{a_x}{a_{\rm re}}\right) \simeq 10^{-9}\,. \tag{4.36}$$

This is comparable although slightly more restrictive than Eq. (4.35). This suggests that there can be a region in parameter space where the second constraint is violated (suggesting $1 \gtrsim f(g,m) \gtrsim 0.1$) but the first satisfied (so that the field still begins to move out of its unstable position). The resulting reheating temperature in this region is also fairly insensitive to changes in g and m, compared with when the second constraint is well satisfied (and the reheat temperature influenced by $f(g,m)$ as given by equation (4.17)). Nevertheless, to ensure that we are in the region of parameter space for which the field has oscillated enough to reheat the universe, when referring to these constraints we will use $(m/m_{\rm pl})^2 g \geq 10^{-8}$. Having $m_{\rm pl}$ as the largest possible mass scale in the model , the strength of the ϕ-χ interaction is therefore bounded from below with $g \geq 10^{-8}$.

A second constraint exists from the requirement that the χ field does not influence the motion of the ϕ field. First we will consider the requirement the negative potential

energy from the coupling term is not significant compared to the kinetic energy of ϕ. As the kinetic energy is constantly diminishing, it will be simpler to overestimate the potential energy and underestimate the kinetic energy. We will therefore take the potential energy to be its maximum value, and the kinetic energy to be the value at $\phi = m$. No matter the evolution of these quantities, if the quantities evaluated at these two time intervals are not comparable, they never will be. Reading off the kinetic energy from equation (4.4) and requiring this always greatly exceeds the maximum of the negative potential energy gives,

$$\lambda m_{\rm pl}^4 \left(\frac{a_x}{a_m}\right)^6 > \frac{m^4 g^4}{4\lambda_\chi}\,, \tag{4.37}$$

or

$$\left(\frac{m}{m_{\rm pl}}\right)^4 \frac{g^4}{\lambda_\chi} < 10^{-19}\,. \tag{4.38}$$

If we choose values to satisfy $\Delta t \gg \tau$,

$$\frac{g^2}{\lambda_\chi} \ll 10^{-3}\,. \tag{4.39}$$

Now consider the effect of the coupling on the motion of ϕ directly. The equation of motion gives

$$\ddot{\phi} + 3H\dot{\phi} - g^2\chi^2\phi = 0\,, \tag{4.40}$$

where the kinetic energy is treated as the dominant contribution to energy of the ϕ field. The third term is the contribution due to the coupling, and we wish to ensure this is negligible compared to the second. Using equation (4.3) to express Hubble's parameter in terms of the kinetic energy $\rho_\phi \approx 1/2\dot{\phi}^2$, this condition can be written as,

$$\sqrt{\frac{4\pi}{3}}\frac{6}{m_{\rm pl}}\rho_\phi \gg g^2\chi^2\phi\,. \tag{4.41}$$

Treating the right-hand side as the average value over oscillations of the χ field, we can rewrite it using $\rho_\chi = g^2\phi^2\langle\chi^2\rangle$. Writing these energy densities out explicitly, including their evolution with scale-factor gives

$$\sqrt{48\pi}\lambda\left(\frac{a_x}{a_m}\right)^6\left(\frac{a_m}{a}\right)^3 \gg f(g,m)\left(\frac{m}{m_{\rm pl}}\right)^3\frac{g^4}{4\lambda_\chi}{\rm e}^{-\Gamma_\chi({\rm t}-{\rm t_m})}\,. \tag{4.42}$$

The right-hand side will quickly become tiny when $t \sim \Gamma_\chi^{-1}$. We need only consider the value of $a \propto t^{1/3}$ when this occurs, and then $(a_m/a)^3 \simeq \Gamma_\chi t_m$. Using $\Gamma_\chi \simeq \alpha g m$ and Eq. (4.17) gives:

$$\left(\frac{m_{\rm pl}}{m}\right)^{2/3}\frac{g^{5/3}}{\lambda_\chi} \ll \sqrt{2}\alpha\lambda^{-1/6}\left(\frac{a_m}{a_x}\right) \simeq 3\times 10^3\alpha\,, \tag{4.43}$$

and thus unless the decay rate is tiny and the reheating period too long, once we fulfill the other constraints this is practically always fulfilled. In other words, having the parameter values such that the field χ performs some oscillations around the minimum while the expansion rate is dominated by the inflaton kinetic energy, those oscillations will not backreact onto the motion of the inflaton field.

4.5 Range of temperatures

The constraints show there can be significant variation in the resulting temperature. The energy density depends most sensitively on g, and the range this parameter can take is severely constrained by the other parameters. If λ_χ is too small, g must be made small enough to avoid interfering with the evolution of ϕ. If m is made too small, g must be made large enough to ensure the field reacts while it is unstable. For instance, if $\lambda_\chi \sim 1$ then g can range from $g \sim 10^{-2}$ at its largest (with $m/m_{\rm pl} \sim 10^{-3}$) to $g \sim 10^{-6}$ (with $m/m_{\rm pl} \sim 10^{-1}$), assuming m is kept below the Planck scale. Decreasing λ_χ can constrain it further.

Reheating ends by the time $t \simeq \Gamma_\chi^{-1}$. Plugging Eqs. (4.31), (4.17), with $\Gamma_\chi \simeq \alpha g m$ and $\lambda = 10^{-14}$, we have:

$$\rho_{\rm r} \simeq 8 \times 10^{-6} m_{\rm pl}^4 \alpha \left(\frac{m}{m_{\rm pl}}\right)^{4/3} \frac{g^{11/3}}{\lambda_\chi} \ln(\Gamma_\chi t_m)^{-1} \,. \tag{4.44}$$

Therefore, working for example with $\lambda_\chi \simeq 1$ and $\alpha \simeq 10^{-4}$, we find the range

$$g_*^{1/4} T_{\rm RH} \sim (10^{11} - 10^{14}) \text{ GeV} \,. \tag{4.45}$$

One can also look at very weak coupling, where for example $g \simeq 10^{-4}$, $\lambda_\chi \simeq 10^{-5}$, $\alpha \simeq 10^{-4}$, and $m/m_{\rm pl} \simeq 10^{-2}$ gives $g_*^{1/4} T_{\rm RH} \simeq 3 \times 10^{13}$GeV. In Fig. 4.3 we give the range for the reheating temperature $T_{\rm RH}$ in the plane $g - m/m_{\rm pl}$ for two values of λ_χ. As we decrease the value of the χ self-coupling, the allowed region in the plane will be further reduced, and that would be the main effect on $T_{\rm RH}$.

4.6 Transition to radiation domination

The kinetic energy of the ϕ field is dropping quickly. Peebles and Vilenkin noted that unless radiation domination occurred before the kinetic energy reached the potential energy in ϕ, the universe would return to an inflationary regime from which it would

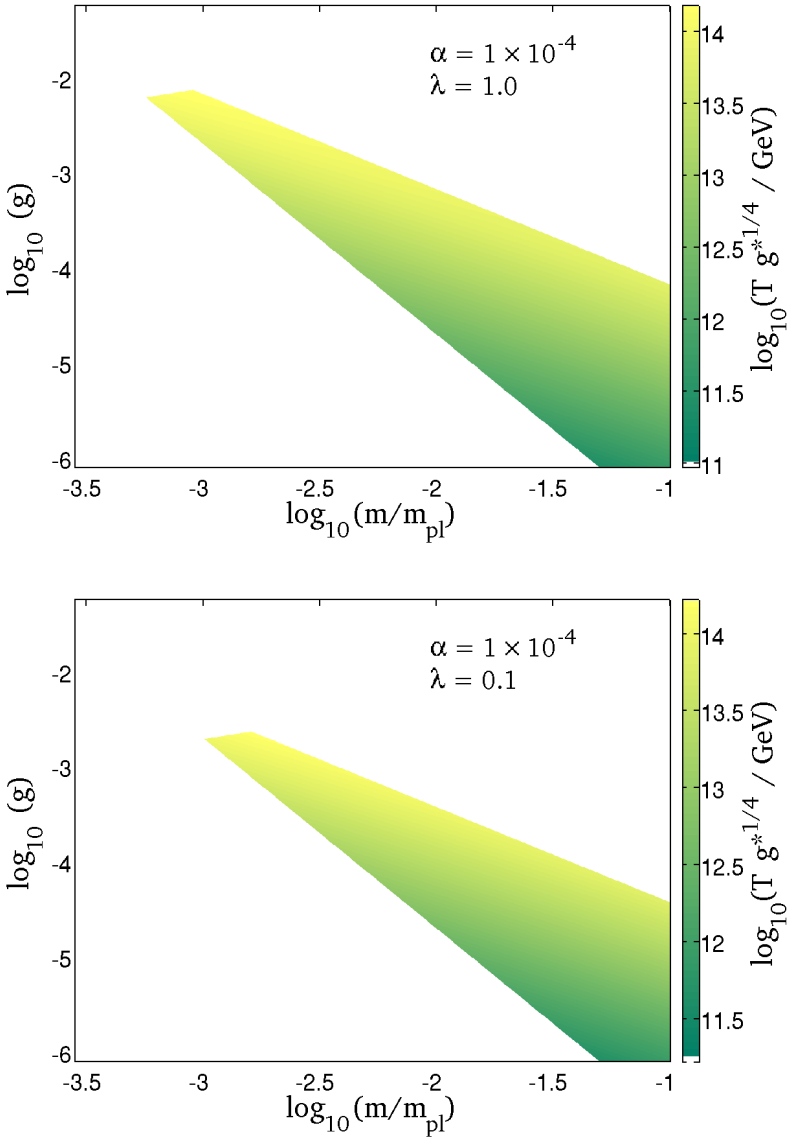

Figure 4.3: The range for the reheating temperature $T_{\rm RH}$ in the plane $g - m/m_{\rm pl}$, for $\lambda_\chi = 1$ (top panel) and $\lambda_\chi = 0.1$ (bottom panel). Decreasing the value of λ_χ has the effect of reducing the allowed region of parameter space. The shading density indicates high (light) to low (dark) reheating temperature in the allowed region of parameter space.

never recover. Our model has the same requirement. We will write $a_{\rm end}$ as the scale-factor when reheating ends ($t \sim \Gamma_\chi^{-1}$) and the radiation redshifts as an unsourced relativistic fluid.

The kinetic energy approaches the potential energy at a scale-factor a_*, given by

[115]

$$\frac{a_*}{a_{\rm end}} < \frac{a_x}{a_{\rm end}} \left(\frac{m_{\rm pl}}{M}\right)^{4/3} \ln(m_{\rm pl}/M)\,. \tag{4.46}$$

The energy in the ϕ field is

$$\rho_\phi \simeq \lambda m_{\rm pl}^4 (\Gamma_\chi t_m)^2 \left(\frac{a_x}{a_{\rm end}}\right)^6 \left(\frac{a_{\rm end}}{a}\right)^6 , \tag{4.47}$$

and the energy in the radiation is given in Eq. (4.44). The ratio is of order unity when the kinetic energy reaches the energy in the radiation, at a scale-factor of $a_{\rm r}$. We require $a_{\rm r} < a_*$ to ensure radiation dominated expansion begins. Then,

$$2 \times 10^6 \frac{g^{7/3}}{\lambda_\chi} > \left(\frac{M}{m_{\rm pl}}\right)^{8/3} \ln^{-2}\left(\frac{m_{\rm pl}}{M}\right) , \tag{4.48}$$

which is easily satisfied for the values of the parameters we have been considering, but does place a weak upper bound on λ_χ for a given choice of g.

Importantly, the energy in the potential of ϕ varies only very slowly. The exact scale-factor $a_{\rm end} < a_{\rm r} < a_*$, where radiation transition occurs, matters very little in terms of its future evolution: the variation of the field's potential energy in the range in which this can happen is negligible, as the field is moving so slowly. The same evolution outlined in [115] therefore takes place, with ϕ mimicking a cosmological constant and today's dark energy density obtained with $M \sim 10^6$GeV .

4.7 Domain walls

Typically in models with a symmetry breaking term, domain walls inevitably form via the Kibble mechanism [85]. Parts of the universe causally separated have no way of being correlated. When the χ field becomes unstable, in some regions the field will move to positive values, and in other regions to negative values. The domain walls created by the smooth transition between these values are generally a serious problem in many cosmological models, as they can come to dominant the energy density of the universe [145].

Fortunately, and unlike in many models with such an occurrence, the symmetry is restored once $\phi > m$. Any walls formed will then dissolve. However, this could leave some effect upon the amount of reheating taking place within the regions of the domain wall, potentially influencing large scale structure formation. As the wall thickness is significantly smaller than a horizon, we do not expect this to be a large effect, but it warrants further investigation.

4.8 Summary and future discussion

By introducing an additional coupling to a second, self-interacting scalar field, we can restore the traditional reheating mechanisms to the scenario proposed by Peebles and Vilenkin [115]. The symmetry breaking term leaves the field temporarily unstable, which allows it to gain significant amounts of energy. Using a phenomenological approach to reheating, we have calculated the evolution of the relativistic particles produced by the decay of this field. We then described a variety of constraints on the model parameters, ensuring that the additional field does not interfere with the behaviour of the original inflaton. The range in temperature produced is quite narrow if the new field does not strongly self-interact, and the symmetry breaking scale does not reach the Planck scale. But with a self coupling of the order of $\lambda_\chi \gtrsim 0.01$, one has $g_*^{1/4} T_{\rm RH} \sim 10^{11} - 10^{14}$ GeV. This allows us to suppress the contribution of the gravitational waves at the time of BBN well below the upper limit.

Observationally, we expect no differences in this scenario to the standard ΛCDM cosmology with an inflationary beginning. While inflation with a quartic potential is now severely constrained (if not ruled out) by the latest CMB observations [91], there is no obvious reason the potential cannot be returned to a standard quadratic one in the inflationary era (though the exact details have yet to be worked out). Meanwhile, the quintessence era appears just as a cosmological constant; the scalar field is frozen. Any deviations from $w = -1$ would be extremely difficult to observationally determine.

With a model that tries to explain the evolution of the Universe from early inflation to today's dark energy domination, one of the issues to explain is the origin of the baryon asymmetry [80]. A possibility would be spontaneous baryogenesis through a derivative coupling of the inflaton field to a matter current [49]. In our set-up, due to the larger coupling of χ to the light degrees of freedom (compared with the gravitational couplings), thermalization occurs promptly after the start of reheating. This opens up the possibility for example of having leptogenesis during or after reheating, by the decay of the lightest right handed neutrino. The field χ being quite heavy during reheating could decay into the lightest right handed neutrino, or alternatively the neutrinos could be thermally produced.

The model could potentially form domain walls, often causing problems in similar models. In this variation, we note they are transient and do not interfere with future evolution of the universe. The possibility exists their effect upon the reheating temperature will appear as an imprint on large scale structure formation.

The analyses of the reheating mechanism proposed in this chapter has been done at tree-level. But having coupled the quintessence field to another scalar field, with a strength $g \gtrsim 10^{-6}$, the question arises about the stability of the quintessential potential to quantum corrections. Here we argue that indeed these corrections are under control due to the decoupling theorem, and that the scenario is not spoiled by quantum corrections, but leave the explicit calculation for a future work.

Quantum corrections with the χ field running in the loops can give rise to a potentially large correction to the quintessence potential, with the one-loop correction given by [43]:

$$\Delta V^{(1)} = \frac{1}{64\pi^2} \sum_{i=\phi,\chi} m_i(\phi)^4 \left(\ln \frac{m_i^2(\phi)}{\mu^2} - \frac{1}{2} \right) , \tag{4.49}$$

where μ is the renormalization scale and $m_i(\phi)$ the field dependent masses, with $m_\chi = g\phi$. So for the χ field we have a very heavy state $m_\chi \sim g m_{\rm pl} \gg gM$ that is excited in a universe with an energy density $\rho \ll \lambda M^4$. Given that we do not have enough energy to excite such a heavy states, physically we can expect that they decouple from the spectrum [139, 7], and their contribution to the effective potential is highly suppressed. However, the 1-loop effective potential as given in Eq. (4.49) is computed using a mass independent renormalization scheme that does not take into account threshold effects. The decoupling would appear naturally when using instead a mass dependent renormalization scheme [66]. To deal with this problem when working with the effective potential one can use instead the "improved" effective potential [15, 16, 34] by replacing all parameters (masses, couplings and field vevs) by their renormalized values in both the tree level and one loop potential, and imposing the physical condition that the potential does not depend on the renormalization scale, $dV(\mu)/d\ln\mu|_{\mu=\mu^*} = 0$. By choosing the renormalization scale μ^* below any heavy mass threshold in the model, heavy states decouple and the dependence on μ is minimised. We are then left with the tree-level potential again with all the parameters evaluated at μ^*,

$$\begin{aligned} U(\phi,\chi) &= V(\phi, M(\mu^*), \lambda(\mu^*)) \\ &\quad + \frac{g^2(\mu^*)}{2}\chi^2\left(\phi^2 - m^2(\mu^*)\right) + \frac{\lambda_\chi(\mu^*)}{4}\chi^4 . \end{aligned} \tag{4.50}$$

All possible quantum corrections due to the heavy states are then encoded in the running of the mass parameters and couplings through the renormalization group equations, and are therefore expected to be under control. In the following chapter, we will see an explicit calculation of quantum corrections with decoupling in this model.

Chapter 5

Decoupling in Effective Potentials

As we mentioned briefly at the end of the previous chapter, the corrections quantum effects make to the potential $V(\phi)$ of some cosmological scalar field model are important. Both the magnitude of the potential and its steepness are important for a viable cosmological model, and if the corrections are extremely large, this will at best invalidate neglecting quantum corrections entirely (requiring the extremely difficult task of calculating the corrections non-perturbatively), or at worst invalidate the model.

Methods exist for calculating the quantum corrections to the potential, order by order in $\hbar$. The most common approach is to calculate the Coleman-Weinberg effective potential. However, most renormalisation schemes neglect the effect of decoupling - that heavy fields should not influence the low energy behaviour of the theory. This is especially true for models with large field values (such inflation or quintessence), where a coupling can result in an extremely large effective mass of the second field, due to the large field value of the first. Neglecting this effect can result in overly large quantum corrections brought about by the coupling terms, incorrectly limiting the size of coupling parameters allowed.

To demonstrate how including decoupling can reduce the size of quantum corrections (and so increase the allowed size of couplings), we will proceed to discuss how all these calculations proceed, before applying the formalism to models of cosmological scalar fields.

5.1 Quantum field theory

Before proceeding to the main point of this chapter, we will take some time to review the ingredients of quantum field theory. This section is based upon several textbook accounts [113, 125, 118].

5.1.1 Quantisation

It is possible to quantise a classical field theory in the canonical Hamiltonian approach. This is often simpler than more advanced path integral approach. This canonical formalism begins by taking the classical degrees of freedom (in quantum mechanics this is the position q, in field theory this is the field value at each point in spacetime $\phi(x)$) and promoting them to operators ($\hat{q}$ or $\hat{\phi}$, though we shall usually omit the explicit hat notation for operators). The final step is to impose commutation relations with the conjugate momenta p (or conjugate momentum density $\pi(x)$ in the case of a field).

In fact, many of the basic results can be obtained easily from this approach, including carrying out perturbation theory and computing cross-sections. These calculations can involve computing many integrals. Despite their multitude, the sorts of integrals encountered have an organisation to them, and as an aid to calculation the *Feynman rules* were developed. These are rules for constructing *Feynman diagrams*, as well as mapping these diagrams back to the complicated integrals they represent. Rather than deriving the details from scratch each time, computations in QFT often amount to understanding the problem in terms of which diagrams are relevant, and then reassembling the mathematics to be solved.

In the reverse fashion, often perturbative expansions will appear as a series of integrals that can be viewed as an expansion in diagrams. Important results in field theory that might normally be expressed as mathematical statements regarding integrals (or series of integrals) are often stated in terms of diagrams instead, particularly as the permutations of diagrams can be easier to understand than the permutations of a series of explicit integrals, making manipulations easier in the diagrammatic form.

Many of these integrals that have to be evaluated to compute even something as apparently mundane as a decay-rate or a cross-section turn out to formally diverge. This was resolved by 'renormalisation'. The approach we will take here is renormalised perturbation theory, where the original parameters in the Lagrangian are split into a renormalised part, and a 'counter-term' that will be chosen purely to remove the

troubling infinity. We shall see that this has interesting theoretical consequences.

The more modern approach for quantisation uses Feynman's path integral formulation of quantum mechanics as a starting point. The basic argument can be reduced to the notion of taking the well known double slit experiment to its almost absurd extreme, by adding additional screens and additional slits in the screens. Applying the same logic to the simplest version, a superposition of every single possible path between the source and the considered end-point is required to calculate the probability of the particle's end location, appropriately adjusting for phase as the length of each path changes. Taken to this ridiculous extreme of an infinite number of screens with an infinite number of slits, *any* path through space (no matter how classically unusual) must be considered.

Feynman's result for quantum mechanics was that the probability for finding a particle at position q' at time t', given that it was at position q at time t, could be expressed as

$$\langle q't|q,t\rangle \propto \int \mathcal{D}q \exp\left[\frac{i}{\hbar}\int_t^{t'} L(q,\dot{q})\mathrm{d}t\right], \tag{5.1}$$

The symbol $\mathcal{D}q$ is short-hand for integrating over every path the particle might take ;

$$\int \mathcal{D}q \propto \lim_{n\to\infty}\int \prod_{i=1}^{n-1}\mathrm{d}q_i\,. \tag{5.2}$$

This *functional integration* has its counterpart operation, that of *function differentiation*, where in four dimensions the basic axiom is

$$\frac{\delta}{\delta J(x)} = \delta^{(4)}(x-y) \tag{5.3}$$

or

$$\frac{\delta}{\delta J(x)}\int \mathrm{d}^4y J(y)\phi(y) = \phi(x)\,. \tag{5.4}$$

This naturally generalises the rule of discrete vectors to continuous ones,

$$\frac{\partial}{\partial x_i} = \delta_{ij} \tag{5.5}$$

or

$$\frac{\partial}{\partial x_i}\sum_j x_j k_j = k_i\,, \tag{5.6}$$

in the same way that functional integration generalises the notion of an integration of over a variable to the integration over all functions. As a final technical note, a functional translates a function into a number, as opposed to a function which returns

a number when given one (or, a function if given a function, as the function of a function is another function).

Of course this result is consistent with our classical approximations, as the paths are each weighted by e^{iS} (recall the definition of the action S is the integral over time of the Lagrangian), and so oscillate rapidly as the path is changed. If the weight remains stationary with respect to changes in the path, these will add coherently and so contribute the most, so

$$\delta \int L \mathrm{d}t = 0 \,. \tag{5.7}$$

The classical principle of the stationary action is derived as a limit.

In quantum field theory, the main result is very similar,

$$\langle \phi'(\mathbf{x}) | \phi(\mathbf{x}), t \rangle \propto \int \mathcal{D}\phi \exp\left[\frac{i}{\hbar}\int \mathcal{L}\right] \mathrm{d}^4 x \,, \tag{5.8}$$

with the main differences only the move towards a more relativistic appearance, and the field replacing the position as the degree of freedom.

5.1.2 The effective potential

In field theory we know that the vacuum is not quite exact; the existence of particles is ambiguous within the bounds of the uncertainty principle. A way of including this is to include a source $J(x)$ for the field (analogous to how a current sources the electromagnetic field), which acts as a source (or sink) for such events. Finally, to pick out the vacuum state $|\Omega\rangle$, we can either include a small imaginary quadratic term, or equivalently rotate the time direction by an infinitesimal amount (see e.g.[125]). We will not worry too much about this and simply absorb it into our notation, so we are then left with (setting $\hbar = 1$),

$$\langle \Omega, t = \infty | \Omega, t = -\infty \rangle \propto Z[J] \equiv \int \mathcal{D} \exp\left[i \int (\mathcal{L} + J\phi)\, \mathrm{d}^4 x \right] . \tag{5.9}$$

This has a great deal of importance, the main one being that $n-point$ correlation functions $\langle \Omega | T[\phi_1(x)\phi_2 ... \phi_n] | \Omega \rangle$ (the T operator merely orders the contents in order of time coordinate) can be calculated by taking repeated functional derivatives of $Z[J]$. For our purposes we only want the one-point correlation function, which is just the expectation value of the field in a vacuum, given a source J,

$$i \frac{\delta}{\delta J(x)} \log Z = -\langle \Omega | \phi(x) | \Omega \rangle_J \,. \tag{5.10}$$

We can use this to define the notion of a classical field $\phi_{\rm cl}$, about which quantum corrections are assumed to be small. In this sense $\phi_{\rm cl}$ serves as the expectation value of the field,

$$\phi_{\rm cl} = \langle\Omega|\phi(x)|\Omega\rangle_J \,. \tag{5.11}$$

We would like this classical field to obey the classical equations of motion; what action principle does it obey?

If we define $E[J]$ via, $Z[J] = \exp(-iE[J])$, then we see that the functional derivative of $E[J]$ is nothing more than $-\phi_{\rm cl}$. Furthermore, we can carry out a Legendre transformation of $E[J]$

$$\Gamma[\phi_{\rm cl}] \equiv -E[J] - \int \mathrm{d}^4y J(y)\phi_{\rm cl}(y) \,. \tag{5.12}$$

This is just the requirement for

$$\frac{\delta}{\delta\phi_{\rm cl}}\Gamma[\phi_{\rm cl}] = -J(x) \,, \tag{5.13}$$

or in the absence of a source, Γ serves as an effective action for the classical field $\phi_{\rm cl}$. If we can only compute the effective action, we will understand how the classical field (the expectation value of the quantum field) behaves, including quantum mechanical effects.

To compute the effective action, we will follow [118] and split the Lagrangian into a main part and one containing the aforementioned counterterms,

$$\mathcal{L} = \mathcal{L}_1 + \delta\mathcal{L} \,. \tag{5.14}$$

The lowest order, we know that we just have the classical field equations,

$$\left.\frac{\delta\mathcal{L}}{\delta\phi}\right|_{\phi=\phi_{\rm cl}} + J(x) = 0 \,(\text{lowest order}) \,. \tag{5.15}$$

Defining J_1 as whatever satisfies this exactly when $\mathcal{L} = \mathcal{L}_1$, we can treat the difference between J and J_1 like a counterterm,

$$J(x) = J_1(x) + \delta J(x) \,, \tag{5.16}$$

where like all other counterterms, it must be determined as we proceed through each order in perturbation theory to match the original definition of $\phi_{\rm cl}$ (i.e. the vacuum expectation value given a source J).

Using this notation,

$$e^{iE[J]} = \int \mathcal{D}\phi \exp\left[i\int \mathrm{d}^4x \left(\mathcal{L}_1[\phi] + J_1\phi\right)\right] \exp\left[i\int \mathrm{d}^4x \left(\delta_{\mathcal{L}}[\phi] + \delta J\phi\right)\right] \,. \tag{5.17}$$

Setting aside the second exponential (containing only counterterms), the next step is to break ϕ into $\phi(x) = \phi_{\rm cl} + \eta(x)$. The exponent then becomes (restoring $\hbar$ once again for clarity),

$$\begin{aligned}\int \mathrm{d}^4x\,(\mathcal{L}_1 + J_1\phi) &= \int \mathrm{d}^4x\,(\mathcal{L}_1[\phi_{\rm cl}] + J_1\phi_{\rm cl}) && (5.18)\\ &+ \int \mathrm{d}^4x\eta(x)\left(\frac{\delta\mathcal{L}_1}{\delta\phi} + J_1\right) && (5.19)\\ &+ \frac{\hbar}{2}\int \mathrm{d}^4x\,\mathrm{d}^4y\,\eta(x)\eta(y)\frac{\delta^2\mathcal{L}_1}{\delta\phi(x)\delta\phi(y)} + \ldots\,, && (5.20)\end{aligned}$$

where functional derivatives of $\mathcal{L}_1$ are evaluated at $\phi_{\rm cl}$.

The term linear in η vanishes due to the definition of J_1, while cubic and higher terms we will neglect for now - it can be shown these correspond to Feynman diagrams with multiple loops. In fact, this entire expansion is diagrammatically an expansion in diagrams with steadily increasing numbers of loops. The first term contains diagrams with no loops, while the terms proportional to $\hbar$ contain those with single loops. Higher order terms contain diagrams with more loops.

Keeping the term quadratic in η means the integral is a Gaussian integral, and so results in a functional determinant. Taking the log to find $E[J]$, and then carrying out the Legendre transform,

$$\begin{aligned}\Gamma[\phi_{\rm cl}] &= \int \mathrm{d}^4x\mathcal{L}_1[\phi_{\rm cl}] + \frac{i\hbar}{2}\log\det\left[-\frac{\delta^2\mathcal{L}_1}{\delta\phi\delta\phi}\right] && (5.21)\\ &- \text{ diagrams + counter terms}\,. && (5.22)\end{aligned}$$

Thus we see our zeroth order action is merely the classical one; the next term gives the first order correction from quantum mechanics, and neglecting higher order terms leaves only the counter-terms as the final addition.

To evaluate this functional determinant, consider a Lagrangian density for a single scalar field with some potential $V(\phi)$. Then we can evaluate the contents of the bracket, and write the first order correction to the effective action as

$$\Gamma[\phi_{\rm cl}]^{(1)} = \frac{i\hbar}{2}\log\det\left[\partial^2 + V''\right]\,, \tag{5.23}$$

with primes indicating derivatives with respect to ϕ.

Next we use the identity: $\log\det M = Tr\log M$. The trace of an operator is just the sum of its eigenvalues, so we need to sum the eigenvalues of the operator $\log(\partial^2 + V'')$.

In order to sum these eigenvalues up, the following useful result is needed. Consider a set of generalised vectors $|x\rangle$, normalised such that

$$\langle x\,|x'\rangle = \delta^{(d)}(x - x') \tag{5.24}$$

where d is the number of spacetime dimensions. The completeness relation generalises from a discrete space to the continuous space in the following fashion:

$$\hat{I} = \int \mathrm{d}^{\mathrm{d}}x \; |x\rangle\langle x| \,. \tag{5.25}$$

Consider the eigenvalue problem:

$$\hat{F}\phi_n(x) = \lambda_n \phi_n(x). \tag{5.26}$$

The covariant scalar product over a general spacetime with metric determinant g can be defined:

$$(f, g) \equiv \int \mathrm{d}^{\mathrm{d}}x \, \sqrt{g} f(x) g(x) \tag{5.27}$$

If a discrete set of eigenvalues is assumed, then

$$\int \mathrm{d}^{\mathrm{d}}x \, \sqrt{g} \phi_m(x) \phi_n(x) = \delta_{n,m} \tag{5.28}$$

Another way of writing an eigenvalue problem is in Dirac notation. Consider an operator $\hat{O}$ that has identical eigenvalues to $\hat{F}$, namely λ_n.

$$\hat{O} \, |\psi_n\rangle = \lambda_n \, |\psi_n\rangle \tag{5.29}$$

The states are normalised: $\langle \psi_n \, | \psi_m \rangle = \delta_{n,m}$. Now, inserting a factor of unity into this eigenvalue problem,

$$\hat{O}\hat{I} \, |\psi_n\rangle = \int \mathrm{d}^{\mathrm{d}}x' \, \hat{O} \, |x'\rangle\langle x'| \, \psi_n\rangle = \lambda_n \, |\psi_n\rangle \tag{5.30}$$

If we pre-multiply by $\langle x \, |$, then

$$\int \mathrm{d}^{\mathrm{d}}x' \, \langle x \left| \hat{O} \right| x' \rangle \psi_n(x') = \lambda_n \psi_n(x) \tag{5.31}$$

where $\langle x \, | \psi \rangle = \psi(x)$. We can also make the observation

$$\langle \psi_n \, | \psi_m \rangle = \int \mathrm{d}^{\mathrm{d}}x \, \psi_n^\dagger(x) \psi_m(x) = \delta_{n,m}. \tag{5.32}$$

by inserting the identity operator into the normalisation condition. Now compare equation (5.31) with (5.26), and equation (5.32) with (5.28). The eigenvalue problems are identical only if:

$$\psi_n = g^{1/4}(x) \phi_n(x) \tag{5.33}$$

and

$$\langle x \left| \hat{O} \right| x' \rangle = g^{1/4}(x) \hat{F} g^{-1/4}(x) \delta^{(d)}(x - x') \tag{5.34}$$

In flat space-time, this leads to the *useful result*:

$$\langle x \left| \hat{O} \right| x' \rangle = \hat{F} \delta^{(d)}(x - x') \,, \tag{5.35}$$

where $\hat{O}$ and $\hat{F}$ are operators that share the same eigenvalues λ_n. Obviously, they could even be the same operator.

Let us return to the eigenvalue problem, equation (5.31)

$$\int \mathrm{d}^{\mathrm{d}} x' \, \langle x \left| \hat{O} \right| x' \rangle \langle x' \, | \psi_n \rangle \, = \lambda_n \langle x \, | \psi_n \rangle \tag{5.36}$$

Now take $\hat{O}$ to share the same eigenvalues as the operator we are interested in: $\hat{F} = \log(\partial^2 + V'')$. Then apply the useful result to yield:

$$\int \mathrm{d}^{\mathrm{d}} x' \, \log(\partial^2 + V'') \delta^{(d)}(x - x') \langle x' \, | \psi_n \rangle \, = \lambda_n \langle x \, | \psi_n \rangle \tag{5.37}$$

How does our operator act on $\delta^{(d)}(x - x')$? This is most easily addressed in Fourier space, using the identity

$$\delta^{(d)}(x) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}^{\mathrm{d}} \mathrm{k}}{(2\pi)^d} e^{-ik.x}. \tag{5.38}$$

At this point, it is necessary to specialise to a constant ϕ_{cl} so that the operator is invariant under spatial translations. Now, by expanding out the logarithm, acting upon the exponential, and re-summing, we see

$$\log(\partial^2 + V'') \delta^{(d)}(x - x') = \int_{-\infty}^{+\infty} \frac{\mathrm{d}^{\mathrm{d}} \mathrm{k}}{(2\pi)^d} \log(-k^2 + V'') e^{-ik.(x - x')}. \tag{5.39}$$

Because of equation (5.32) we can reduce the right-hand side of our eigenvalue problem to simply the eigenvalue λ_n by multiplying by $\psi_n^\dagger(x)$ and integrating over spacetime coordinates x. Along with the expansion of our operator into k-space, thanks to the useful result, we find:

$$\int \mathrm{d}^{\mathrm{d}} x' \int \mathrm{d}^{\mathrm{d}} x \int \frac{\mathrm{d}^{\mathrm{d}} \mathrm{k}}{(2\pi)^d} \log(-k^2 + V'') e^{-ik.(x - x')} \langle \psi_n \left| x \right\rangle \langle x' \right| \psi_n \rangle = \lambda_n. \tag{5.40}$$

The inner-products on the left-hand side are just numbers - we can reorder them as we choose. So we can write:

$$\langle \psi_n \left| x \right\rangle \langle x' \right| \psi_n \rangle = \langle x' \, | \psi_n \rangle \langle \psi_n | \, x \rangle = | \psi_n \rangle \langle \psi_n | \, \delta^{(d)}(x' - x) \tag{5.41}$$

where the outer-product $|\psi_n\rangle\langle\psi_n|$ has been treated as an operator (which it is), and the useful result applied once more.

The integration over x' is easy; the delta function picks out only the value in the exponential where $x = x'$, giving a factor of unity. The factor of $|\psi_n\rangle\langle\psi_n|$ also yields

unity when summed over n by virtue of this being a complete set of states. Thus, the desired result - the one loop correction to the effective action is

$$\Gamma[\phi_{\rm cl}]^{(1)} = \frac{i\hbar}{2}\sum_n \lambda_n = \frac{i\hbar}{2}\int \mathrm{d}^{\mathrm{d}}x \int \frac{\mathrm{d}^{\mathrm{d}}\mathrm{k}}{(2\pi)^d} \log\left(\frac{-k^2+V''}{k^2}\right) \tag{5.42}$$

where the extra ϕ-independent term has been added to make the dimensions of the log correct (identical to adding a constant to the overall result). Note that the form of the effective action can be easily related to an effective potential,

$$V_{\rm eff} = V_1 + i\hbar/2 \int \frac{\mathrm{d}^d k}{(2\pi)^d} \log\left(\frac{-k^2+V''}{k^2}\right) + \text{ counter terms}\,. \tag{5.43}$$

This gives the Coleman-Weinberg effective potential; note that the integral formally diverges. The final stages in the calculation are to regularise the integral (which can be done, for example, by imposing a cut-off scale on the integral, and sending the cut-off to infinity) and then choosing appropriate renormalisation conditions to resolve any remaining infinite terms. We will delay this for now; after discussing the implications of quantum corrections in a cosmological context, we will return to these subtleties, and this will occupy much of the remainder of this Chapter.

5.2 Quantum corrections in cosmology

Often in cosmology, scalar fields with appropriate potentials and interactions are studied in order to explain different physical effects. Very well known examples are inflation models, which provide a solution for the horizon and flatness problems of the standard cosmology, and generate a nearly scale invariant primordial density perturbation, which has been tested by the observations of the cosmic microwave background radiation spectrum [54]. The simpler examples are large field models of inflation with a renormalizable potential, a mass term plus a quartic interaction [96]. When the quartic self-interaction dominates, the WMAP normalization of the primordial spectrum demands this coupling to be tiny, $\lambda \simeq 10^{-14}$. Present observational data also indicate that we live in an accelerated expanding Universe today [117, 124], and that around 70% of the total energy density is made of a component, called dark energy, with negative equation of state w, close to -1 [140, 54]. This could be explained by a light rolling scalar field, usually called quintessence [114, 152, 30, 46, 115]. The dynamics of the field is such that at early times, during matter or radiation domination, its energy density ρ_{DE} is subdominant, and it is only today, when the field finds itself evolving in a roughly constant potential, that ρ_{DE} dominates. Commonly, quintessence potentials are given by

either exponentials or inverse powers of the field, i.e, by a non-normalizable potential, which yields naturally a small classical mass of the quintessence field, below the Hubble parameter, without the need of fine-tuning.

In trying to explain the origin of the inflation and quintessence potential from a particle physics model, it is unlikely that it will appear as an isolated entity with no interactions to any other degree of freedom. For example, inflation should be followed by a reheating period, such that we recover a radiation dominated universe at the end of it. The standard picture is that the inflaton couples to other light degrees of freedom into which it decays during reheating [3, 52, 1, 86]. Moreover in warm inflation dynamics [23] interaction of the inflaton field with other fields is needed to produce radiation concurrently with inflationary expansion. For models where the inflaton has sizable coupling to other fields, in order to maintain the required flatness of the potential, supersymmetry is typically used to control large quantum corrections [102]. For quintessence models, there are viable models coupled to the dark matter fluid [59, 69, 81]. Or quintessence fields can directly be coupled through standard renormalizable interactions to other bosons and fermions, providing for example time-varying masses for these degrees of freedom [31, 27, 138, 20, 71]. But once coupled to other species, one should check the stability of the classical results against quantum corrections.

These sources of quantum corrections have been already studied in the literature. For inflation for fields interacting sizably with the inflaton, often these effects are not a problem, but indeed they can either help with the inflationary trajectory, as in supersymmetric hybrid model [56], or lead to different observational signatures [131, 130]. For quintessence, in Ref. [26] the 1-loop corrections due to the self-interactions, among other cases, were studied by regularizing the theory with a cut-off Λ, and they conclude that for reasonable values of the high-energy cut-off, $\Lambda < m_P$, quantum corrections do not spoil the quintessence potential. At early times, when the quintessence energy density is still subdominant, quantum corrections can be quite large compared to the tree-level potential, but still subdominant compared to the other components of the energy density. By the time the quintessence epoch starts, quantum corrections have already become negligible. A similar conclusion is reached in Ref. [65] for the self-couplings. More severe are the constraints on renormalizable couplings to fermions and scalars, which have to be really tiny and negligible to avoid distortion of the quintessence potential [65, 53, 8].

At first glance it seems that the previous studies on quantum corrections to the

quintessence/inflation potential practically forbid couplings to scalars and/or fermions. Here we want to argue against these conclusions for the particular case of quantum corrections due to heavy states, and show that (at 1-loop and 2-loop orders) these corrections can be kept under control once the decoupling theorem is taken into account [139, 7]. We do not address the question of radiative corrections in a general framework, but focus on scalar models with a well defined hierarchy of mass scales. And we do not address either the question of why for example the self-coupling in the inflaton potential or the quintessence mass are tiny, but given a model that can accomplish that, focus on the problem of what happens when coupling such a light state to a much heavier one. How large has to be the hierarchy between the mass scales, and the energy scale range for which the model is stable against radiative corrections becomes a quantitative question that can be answered with the tools presented in this chapter. The result does not depend on whether we study a quintessence or inflation model. As such we will deal with a generic scalar potential $V(\phi)$, where the field has a non vanishing background field value. We will focus on the interactions to scalars, by introducing another scalar χ with potential:

$$V(\chi) = \frac{1}{2}m_\chi^2\chi^2 + \frac{\lambda_\chi}{4!}\chi^4 + \frac{1}{2}g^2\chi^2\phi^2\,. \tag{5.44}$$

The potentially dangerous term is $g^2\phi^2\chi^2$, which can give rise to a large quantum contribution to the effective potential. This coupling induces a large field dependent mass for the χ field, $m_\chi \sim g\phi$, and when $\phi \sim m_P$, which is often the magnitude during inflation [96, 102, 98, 3, 97] or quintessence [114, 152, 30], this can be larger than $\rho^{1/4}$, ρ being the total energy density. However, if we do not have enough energy to excite such heavy states, physically we expect them to decouple from the spectrum [139, 7], and their contribution to the effective potential to be highly suppressed. This has been explicitly proved in [14, 68], in the context of an effective quantum field theory of two scalar fields. At scales below the mass of the heavy scalar, the theory is well approximated by an effective theory of the light particle alone. Integrating out the heavy mass states, their effects are absorbed into the parameters of the theory with the light particle only, plus corrections suppressed as the inverse powers of the heavy mass [25]. The more disparate are the scales for the low-energy and the heavy mass, the smaller the corrections from the heavy states. These proofs of decoupling are for a general effective action. Expanding the effective action in momentum space, the first term at zero momentum gives the effective potential, and therefore these proofs are equally applicable to the problem of interest here.

The effective field theory (EFT) description together with the exact renormalization

group techniques provide the tools to describe physics at different scales [155, 156, 119, 13, 29, 120, 107], and from them the decoupling theorem follows quite naturally. Following the Wilsonian approach, a theory is defined and valid only up to some cutoff or naturalness scale, Λ_0, while physical observables are given at the scale $\Lambda_R < \Lambda_0$ at which experiments are performed. The low-energy action is obtained step by step by integrating out fluctuations with momentum larger than Λ, with $\Lambda_R < \Lambda < \Lambda_0$. As we lower the cutoff Λ, the scaling of the effective action is controlled by the renormalization group flow. The renormalization group equations are the mathematical statement that physics at low-energy (or similarly at the renormalization scale) cannot depend on the high energy cutoff. The renormalization group flow has an infrared attractor, and indeed, as we scale down from high energy to low-energy, the effective action converges toward a finite subset in the space of possible actions [119, 13]. The operators in the effective action can be classified depending on their mass dimensions as irrelevant, with negative mass dimension, and relevant, with positive mass dimension (recall in the standard renormalization group approach, irrelevant operators correspond to non-renormalizable operators, while relevant operators are the renormalizable ones). By lowering the cutoff smoothly from Λ to Λ_R, the renormalization group equations dictate how the relevant couplings are rescaled, while the irrelevant couplings are suppressed by powers of Λ_R/Λ; that is, they become more and more "irrelevant" at low-energy. Therefore, unless we perform very precise experiments, in practice only a few of the irrelevant operators have to be taken into account in the calculations. The effects of new physics that may be present at the high energy Λ_0, given in terms of the irrelevant operators, decouple from the low-energy action described by the relevant ones at Λ_R, and the low-energy effective action is fully predicted. One can then identify the scale of new physics with that of a heavy massive state. The effective field theory description leads therefore to the decoupling of the heavy states from the low-energy physics. Decoupling here means power suppression by the light to heavy mass ratio: the larger the separation between the scales, the more negligible the heavy mass effects.

The renormalization group equations (RGE) obtained in the EFT description are equivalent to the standard renormalization scheme in quantum field theory (QFT). In both the renormalized parameters are obtained by demanding the effective action to be independent of the high energy cut-off Λ_0. The original parameters of the theory at Λ_0 are the 'bare' parameters, which are unobservable [14, 68, 156, 155, 119, 13, 29, 120, 107]. Physical observables are given in terms of the renormalized parameters, with a scale dependence given by the RGEs. However, in

the standard renormalization procedure the cut-off is sent to infinity, while in the EFT approach the high energy cut-off is finite and sets the scale of validity of the theory, the scale at which new physics may appear. But with or without sending the cut-off to infinity, renormalization requires first a regulator to compute the cut-off dependence of the loop corrections, and a set of normalization conditions to define the renormalized couplings. Different renormalization schemes leads to different relations among the renormalized parameters and the physical observables, but obviously to the same physics. One can choose a renormalization prescription because of mathematical convenience such that the calculations are easier to deal with, which is usually the case of the Minimal Subtraction ($\overline{\text{MS}}$) scheme; on the other hand, the choice can be dictated by the physical problem to be addressed. In our case we want to study the effects of heavy massive states during the cosmological evolution of a light scalar field, that being the inflaton or the quintessence field. This evolution is governed by the effective potential, which we will compute following the standard perturbative renormalization techniques, and our aim is to show the decoupling of the heavy states. To properly address this issue, one must then compute the effective potential using renormalized perturbation theory in which decoupling is already implemented. This is a well studied problem [16, 112, 34], and we want to examine the implications of their results in the context of scalar field inflation and quintessence.

Still, it is clear that physics depends on the scale at which it is observed, and from the EFT approach (or the more standard renormalization procedure) this means that the parameters in the effective potential have to be specified at some renormalization scale. However there is no direct information in the potential about how to choose that scale. A large value of the potential, or its curvature, etc..., has no direct relation with the choice of the renormalization scale. One needs outside information on the model, either a set of physical observables or physical conditions at a given scale to be related to the renormalization parameters. For example, when computing the Higgs effective potential one does not know the self coupling parameter at the electroweak scale. One procedure then is to invoke a higher symmetry, like a Grand Unification Theory (GUT) model, which then specifies the coupling at the GUT scale, and use the renormalization group equations (RGE) to run it down. This is the standard procedure when studying the effective potential in both the Standard Model and the Minimal Supersymmetric Standard Model. If there is no theoretical argument from the symmetries of the model, another possibility would be to do some scattering experiment from which to extract the value of the self-coupling at some momentum scale, similar to what is done for

gauge couplings [6]. Then, repeating the experiment at different energy scales one can further confirm the predicted running [6].

In inflation and quintessence models most often they are not embedded in a higher theory, and moreover outside phenomenological information about the parameters is unavailable. For example, in chaotic inflation normalizing the potential to the cosmic microwave background (CMB) amplitude leads to a tiny ϕ self-coupling of $\lambda_\phi \simeq 10^{-14}$ [54, 96]. This was all done through a classical calculation. What renormalization scale has this specification of $\lambda_\phi \simeq 10^{-14}$ been done at? No information internal to this calculation tell us about the scale. This is a common problem in this sort of inflationary model building. As we will discuss in this chapter, this implies an arbitrariness in the model predictions.

In the case of our model, there are three different couplings that need to be specified, the ϕ self-coupling λ_ϕ, the χ self-coupling λ_χ, and the coupling g between the two scalar fields. If we could scatter particles, that would give the couplings and masses at that scale, but that possibility is not available. Indeed, this problem is general to any scalar field inflation and/or quintessence calculation. One has no *a priori* information about the values of the couplings at any given renormalization scale, so it is a matter of choice, which leads to considerable ambiguity in the predictions from any given model. To highlight this point about the ambiguity, decoupling has an interesting implication. If one chooses all the parameters at a scale well below all thresholds in the model, then there will be no quantum correction to the effective potential as such, and thus the tree-level potential is the full effective potential. This is the implication of decoupling in application to the effective potential in Refs. [16, 112, 34]. On the other hand, if there are massless states, or one only knows the value of the parameters at some scale intermediate to the masses in the theory, then one must implement the renormalization group improved potential.

In our model, the only thing we know is that either for inflation or quintessence λ_ϕ must be very tiny, although λ_χ and g are unconstrained. We also know that the χ mass is heavy when the inflaton background field value is large, whereas the inflaton mass must be smaller than the Hubble scale, or in the case of quintessence field, it must be very tiny. Following the above logic of a scattering experiment, the energy scale certainly will be below the Planck mass, and an internal consistency for our model would require λ_ϕ to be very tiny. Then decoupling requires that heavy states yield no radiative corrections. In the language of EFT this requires a large hierarchy between the scale at which physics is observed, and the heavy mass scale, such that irrelevant

couplings at low-energy yield practically no effect on the light states parameters and can be safely ignored. How large this hierarchy has to be is then a quantitative question that could be in principle addressed using the RGE tools.

In this chapter the decoupling concept will be developed and applied to inflation and quintessence models using standard perturbative renormalization techniques. Our set-up, is that of a scalar theory with a light state, the inflaton or quintessence, coupled to a heavy scalar. The mass of the latter is controlled by the vev of the light scalar field such that it is the large value of the vev which renders the extra scalar heavy.

Before proceeding, to further help focus on the objectives of this chapter, it is worth mentioning, by way of contrast, how the problem studied here, where decoupling applies, differs from another important example involving heavy mass states, where decoupling does not apply. In particular, in theories with spontaneous symmetry breaking, when the heavy mass is given by the vacuum expectation value (vev) of some other field times some dimensionless coupling, we have decoupling by increasing the vev, but not necessarily when increasing the dimensionless coupling and keeping the vev fixed [84, 129, 14, 68]. In the latter case, the theory will contain a strongly interacting sector, and the light sector becomes apparently non-renormalizable, which indicates its sensitivity to the physics of the heavy sector. This is for example the case of the top quark in the Standard Model and its extensions [120]. On the other hand, by increasing the vev but keeping the coupling perturbative, which is the case for the problem considered in this chapter, the light sector remains renormalizable and one has decoupling.

It is also worth clarifying here what this chapter is and is not attempting to do. We are not attempting to explain why the cosmological constant is small nor any such questions addressing fundamental questions about solving the cosmological constant. Similarly for the mass and self-coupling values of the light field, we do not try to explain their origin. All this chapter is doing is making one technical observation, that for those fields coupled to the inflaton or quintessence scalar field which get very large masses as this scalar field amplitude gets large, the quantum corrections from these fields on the effective potential are power suppressed based on the decoupling theorem [139, 7]. The inflaton/quintessence model is taken as an effective theory valid in a range of energies. Therefore, the question of how small or large can the couplings of the massive state to the light one be can be translated into the question of how large can the range of validity of the effective theory be.

The chapter is organized as follows. In section 5.3 we will first discuss quantum

corrections due to renormalizable interactions to other scalar field, computing the RG improved effective potential at one-loop in a Mass Dependent Renormalization (MDR) scheme. This allows us to take into account in the definition of the renormalization parameters and their running the effects of massive states, and therefore decoupling. Details about the calculation of the RGEs at 1-loop in the MDR scheme are provided in Appendix A, Appendix B gives some 2-loop results, and the 1-loop effective potential is given in Appendix C. In section 5.4 we discuss the non-renormalizable self-interactions relevant for a quintessence field. In order to check the stability of the potential against quantum corrections, we study in section 5.5 an example for an inflation model, and a quintessence potential. We present the summary in section 5.6.

5.3 1-loop corrections: renormalizable interactions

In order to proceed, we will first check the 1-loop corrections due to the renormalizable interactions of an extra χ field coupled to the scalar field ϕ. We assume that ϕ is a light field with a non-vanishing vev, whereas the χ state is heavy (heavier than the Hubble expansion rate), with zero vev. The tree-level potential is given by:

$$V^{(0)}(\phi,\chi) = \Omega + \frac{1}{2}m_\phi^2\phi^2 + \frac{\lambda_\phi}{4!}\phi^4 + \frac{1}{2}m_\chi^2\chi^2 + \frac{\lambda_\chi}{4!}\chi^4 + \frac{1}{2}g^2\chi^2\phi^2\,, \tag{5.45}$$

Loop corrections can give rise to a cosmological constant term Ω, a quartic self-interaction λ_ϕ and mass term m_ϕ^2, and for consistency we include them already at tree-level. The aim is to show that those interactions do not pick up large corrections due to the heavy field, and they can remain as small as required during the inflationary or quintessence phase. We do not couple the ϕ directly to fermions for simplicity, and focus only on scalar couplings, but we allow a Yukawa coupling of the χ field to N_F massless Dirac fermions,

$$\mathcal{L}_{\rm Yuk} = -h\chi\bar{\psi}\psi. \tag{5.46}$$

The immediate problem in computing radiative corrections to the effective potential is that the latter must be renormalization scale independent. However, computing the effective potential order by order in perturbation theory, the renormalization scale μ explicitly shows up, and the point is how to choose that scale. One common approach is such that all log corrections due to different mass scales are kept small and can be resumed. For models with a single mass scale, the perturbative quantum corrections lead to logarithmic terms of the form $\ln^n(m^2/\mu^2)$ and the standard procedure for controlling the large log-terms is to choose the renormalization scale near the mass scale

$\mu \sim m$. For a multi-mass case, such as in Eq. (5.44), logarithmic corrections terms will arise of the form $\sim \ln^n(m_\phi^2/\mu^2)\ln^m(m_\chi^2/\mu^2)$ and if m_ϕ and m_χ are at very different scales, say $m_\chi \gg m_\phi$, there is no ideal choice of μ to control the large logs. However such terms arise in a mass-independent renormalization scheme, which is problematic for multi-mass cases for which there are disparate mass scales, since accounting for decoupling effects can not be done. It is physically better motivated to use a MDR scheme [66], in which any field with a mass much bigger than others in the system has its quantum effects suppressed in powers of the light-to-heavy mass ratio. This is the direct implication of the decoupling theorem proved in the EFT context. Thus in a mass dependent scheme, quantum corrections lead to terms in the perturbative expansion moderated by power suppression, $\sim (\mu/m_\chi)^k \ln^n(m_\phi^2/\mu^2)\ln^m(m_\chi^2/\mu^2)$. In this case the choice of renormalization scale $\mu^2 \ll m_\chi^2$, although will lead to large log terms, it is not important since the power suppression term dominates the perturbation series and keeps it under control.

Resummation of the logs is done by applying renormalization group (RG) techniques to obtain the RG-improved effective potential [83, 82, 15, 63, 40, 57, 62, 61]. For multimass scale problems, the prescription given in Refs. [16, 34] is to always choose the renormalization scale to be of order the lowest mass scale, because the decoupling theorem will ensure that the heavy mass states do not contribute. This can be seen directly when using a mass dependent renormalization (MDR) scheme, because at any order the logs are modulated by the appropriate threshold functions [66, 112]. Recall that these threshold functions suppress those massive contributions that are much above the renormalization scale. Therefore, in the MDR scheme one can immediately see that one gets the most rapid convergence of the perturbative expansion by choosing the RG scale below all thresholds.

In the following we will revise in detail the above arguments, given the calculation of the effective potential, and the resummation of the logs through the RG-improved potential, first by using a mass independent scheme as in [83, 82, 15, 63, 40, 57, 62, 61]. Decoupling of heavy states has to be introduced by hand, by matching the renormalized parameters of the effective potential at high energies (in the sense that all states heavy and light contribute), to the ones in the effective potential at low energies where only light degrees of freedom should appear. We will then compute the effective potential in the MDR scheme, which allows a smooth transition from 'high' to 'low' energies, and directly incorporates the decoupling of heavy states without further assumptions.

1-loop corrections are given by the Coleman-Weinberg potential [43, 149]:

$$\Delta V^{(1)} = \frac{1}{2}\int \frac{d^4q}{(2\pi)^4}\sum_\alpha \ln(q^2+M_\alpha^2)\,, \tag{5.47}$$

where $\alpha = \phi$, χ, and M_α are the field-dependent masses:

$$M_\chi^2 = g^2\phi^2 + m_\chi^2\,, \tag{5.48}$$

$$M_\phi^2 = \frac{\lambda_\phi}{2}\phi^2 + m_\phi^2\,, \tag{5.49}$$

(We will use ϕ for both the quantum field as in Eq. (2) and in all the Lagrangians in this chapter, as well as for the background field value as in the above expressions for M_χ and M_ϕ, since the correct usage will be obvious in each case.) The divergent integrals in Eq. (5.47) can be regularized by using a cut-off Λ, and keeping only terms that do not vanish when the cut-off goes to infinity we have:

$$\Delta V_{reg}^{(1)} = \sum_\alpha \left(\frac{M_\alpha^2}{32\pi^2}\Lambda^2 + \frac{M_\alpha^4}{64\pi^2}\left(\ln\frac{M_\alpha^2}{\Lambda^2} - \frac{1}{2}\right)\right)\,, \tag{5.50}$$

For renormalizable tree-level potentials, the cut-off divergent terms are subtracted by adding counterterms:

$$V_{ct}(\phi) = \delta\Omega + \frac{1}{2}\delta m_\phi^2\phi^2 + \frac{\delta\lambda_\phi}{4!}\phi^4 \tag{5.51}$$

and imposing suitable renormalization conditions at some arbitrary scale μ on the effective potential [43, 149]. This allows to remove the quadratic and logarithmic divergent term, by choosing:

$$\delta\Omega = -\frac{\Lambda^2}{32\pi^2}(m_\phi^2+m_\chi^2) + \frac{1}{64\pi^2}(m_\phi^4+m_\chi^4)\left(\ln\frac{\Lambda^2}{\mu^2} - 1\right)\,, \tag{5.52}$$

$$\delta m_\phi^2 = -\frac{\Lambda^2}{32\pi^2}(\lambda_\phi + 2g^2) + \frac{1}{32\pi^2}(\lambda_\phi m_\phi^2 + 2g^2 m_\chi^2)\left(\ln\frac{\Lambda^2}{\mu^2} - 1\right)\,, \tag{5.53}$$

$$\delta\lambda_\phi = \frac{1}{32\pi^2}(3\lambda_\phi^2 + 12g^4)\left(\ln\frac{\Lambda^2}{\mu^2} - 1\right)\,. \tag{5.54}$$

Adding V_{ct} to Eq. (5.50), the divergent terms cancel out and one is left with the logarithmic contributions depending now on the renormalization scale μ:

$$\Delta V^{(1)} = \frac{1}{64\pi^2}\sum_\alpha M_\alpha^4\left(\ln\frac{M_\alpha^2}{\mu^2} - \frac{3}{2}\right)\,. \tag{5.55}$$

Notice however that there is an arbitrariness in choosing the finite terms in the counterterms, and different renormalization conditions lead to different finite contributions [122] also in the effective potential. The above prescription has been chosen in order to match the standard result for the 1-loop effective potential using dimensional regularization and minimal subtraction ($\overline{\rm MS}$) as a renormalization prescription.

Independently of the implemented renormalization scheme, physics cannot depend on the arbitrary renormalization scale μ, and the effective potential $V = V^{(0)} + \Delta V^{(1)}$ has to satisfy the renormalization group equation (RGE):

$$\mathcal{D}V = \left(\mu\frac{\partial}{\partial\mu} + \beta_{\lambda_a}\frac{\partial}{\partial\lambda_a} - \gamma_\phi\phi\frac{\partial}{\partial\phi} - \gamma_\chi\chi\frac{\partial}{\partial\chi}\right)V = 0\,, \tag{5.56}$$

where λ_a denotes both renormalizable couplings and mass parameters in the potential, $\beta_a = \partial\lambda_a/\partial\ln\mu$ their beta functions, and γ_ϕ, γ_χ are the anomalous dimensions of the fields[1].

The solution to the RGE then provides the RG-improved effective potential [83, 82, 15, 16, 63, 40, 57, 62, 61], given by:

$$V(\phi,\chi,\lambda_a;\mu) = V(\phi(t),\chi(t),\lambda_a(t);e^t\mu)\,. \tag{5.57}$$

We can now evaluate the effective potential at any given scale t by appropriately changing fields, couplings and masses: $\phi(t)$, $\chi(t)$, $\lambda(t)$ are now running parameters, with scale dependence (t-dependence) given by the corresponding RGEs. As remarked in [15], order by order in perturbation theory, the Lth-to-leading log order RGE improved effective potential is given by the $(L+1)$-loop RGE functions, and the L-loop effective potential at some boundary value of t. The main idea of the RG improved method is that by choosing adequately t, the potentially large logs appearing on the LHS of Eq. (5.57) can be resummed. Thus, at lowest order the RG improved effective potential reduces to the tree-level potential with couplings, masses and fields given by the 1-loop running parameters:

$$\begin{aligned} V^{eff} &= \Omega(t) + \frac{1}{2}m_\phi^2(t)\phi^2(t) + \frac{\lambda_\phi(t)}{4!}\phi^4(t) \\ &\quad + \frac{1}{2}m_\chi^2(t)\chi^2(t) + \frac{\lambda_\chi(t)}{4!}\chi^4(t) + \frac{1}{2}g^2(t)\chi^2(t)\phi^2(t)\Big|_{t=t_*}\,. \end{aligned} \tag{5.58}$$

The key point is how to choose t_*, i.e, which is the best choice to evaluate the effective potential. As stressed in Ref. [34], one can choose either a value at which the 1-loop potential has the least μ-dependence, i.e.

$$D(V^{(0)} + \Delta V^{(1)})\Big|_{(t_*)} = 0\,, \tag{5.59}$$

or the scale at which the loop expansion has the best apparent behavior, i.e., $\Delta V^{(1)}(t_*) = 0$. The optimal situation occurs when both criteria are met by the

[1]The anomalous dimensions are the logarithmic μ derivatives of the wave-function renormalization constants of the fields Z_{ϕ_α}. We have included in Eq. (5.56) the contribution of γ_χ for the sake of generality, but we only consider situations when $\chi = 0$, χ referring to the vev of the field.

same choice t_*. This can be done in a model with a single mass scale, say a scalar field ϕ with self-interactions and no couplings to other fields, so that the choice [83, 82, 15, 40, 57, 62, 61]:

$$t_* = \ln \frac{\mu_*}{\mu} = \frac{1}{2} \ln \frac{M_\phi^2(t)}{\mu^2} \,, \tag{5.60}$$

fulfills both conditions, and is equivalent to evaluating the 1-loop potential at the scale $\mu = M_\phi$. But in the presence of very different mass scales, say $M_\phi \ll M_\chi$, the choice is not obvious. The problem is how to rearrange the loop expansion in terms of small parameters for the series expansion to make sense (and to be resummed). With two different mass scales, the correction to the effective potential can be written as [16, 112]:

$$\Delta V = g^2\phi^4 \sum_{l=L-(i+j)}^{\infty} \left(\frac{g^2}{16\pi^2}\right)^l \sum_{i,j}^{\infty} F_{i,j}\left[M_\chi^2/M_\phi^2, \lambda_\phi/g^2\right] s_\chi^i s_\phi^j \,, \tag{5.61}$$

where

$$s_\alpha = \frac{g^2}{16\pi^2} \ln \frac{M_\alpha^2}{\mu^2} \,. \tag{5.62}$$

By taking as a boundary condition either $s_\phi = 0$ or $s_\chi = 0$, to evaluate the L-loop effective potential at given order, there are still potentially large contributions due to $\ln M_\chi^2/M_\phi^2$ in Eq. (5.61), and the series cannot be truncated at any order.

However, as remarked before, one does not (and should not) expect heavy states to modify the low-energy physics, and thus the main issue to address is how to incorporate decoupling of heavy states in the improved effective potential [16, 112, 34], that is, how to get rid of the troublesome large logs by adopting a physical condition. For example, in Eq. (5.58) the decoupling can be incorporated in the running parameters through their RGEs. In a mass-independent renormalization scheme, with no reference to mass scales, decoupling has to be implement by hand by the use of step functions in the RGEs and matching conditions for masses and couplings at each threshold: for a given state, when its mass becomes heavier than the renormalization scale, its contribution drops from the RGEs and the effective potential. In particular we will have [34]:

$$\Delta V^{(1)} = \sum_\alpha \theta(\mu^2 - M_\alpha^2) V_\alpha^{(1)} \,, \tag{5.63}$$

$$V_\alpha^{(1)} = \frac{M_\alpha^4}{64\pi^2}\left(\ln \frac{M_\alpha^2}{\mu^2} - \frac{3}{2}\right) \,, \tag{5.64}$$

and the optimal choice for t_* (or μ_*) is then given by the lower threshold in the model. Indeed in that case all massive states are decoupled, so that $\Delta V^{(1)}(\mu_*) = 0$, i.e., the effective potential is given by the tree-level potential with masses and couplings

evaluated at low-energy. The effect of the heavy states appear when integrating the RGEs from high energy down to the low-energy regime.

However, threshold effects are more naturally taken into account when adopting instead a mass dependent renormalization (MDR) scheme [112]. Following Ref. [66], effective couplings and masses are defined after subtracting the divergences (regularized using dimensional regularization for example) of the 1PI Green functions, by imposing suitable normalization conditions at the euclidean external momentum $p_E^2 = -p^2 = \mu^2$, with μ^2 being the arbitrary renormalization scale. The Appendices give a detailed account of the MDR scheme, beta functions and 1-loop correction to the effective potential in this scheme, while the following just summarizes the main results.

We have already mentioned the scheme-dependence of the finite contributions in the counterterms, or equivalently the renormalization constants Z_a [122]:

$$\begin{aligned} \delta m_\alpha^2 &= m_\alpha^2 (Z_{m_\alpha^2}^{-1} - 1)\,, && (5.65)\\ \delta\lambda_\alpha &= \lambda_\alpha (Z_{\lambda_\alpha} - 1)\,, && (5.66)\\ \delta\Omega &= \Omega (Z_\Omega^{-1} - 1)\,, && (5.67) \end{aligned}$$

that relate the bare parameters (denote by a subindex "0") with the renormalized ones:

$$\begin{aligned} m_\alpha^2(\mu) &= Z_\phi Z_{m_\alpha^2} m_{\alpha 0}^2\,, && (5.68)\\ \lambda_\alpha(\mu) &= Z_\phi^2 Z_{\lambda_\alpha}^{-1} \lambda_{\alpha 0}\,, && (5.69)\\ \Omega(\mu) &= Z_\Omega \Omega_0\,. && (5.70) \end{aligned}$$

By modifying the subtraction conditions, we are explicitly including finite contributions from the 1PI functions, which carry the dependence on the different mass scales of the model. When taking the derivative with respect to the arbitrary renormalization scale μ, that dependence appears in the beta functions as threshold functions depending on the different ratios M_α^2/μ^2. These threshold functions modulate the contribution of each massive state to the running of the different parameters. And in the exact decoupling limit $\mu = 0$ the threshold functions vanish. For the renormalizable parameters in Eq.

(5.45), the beta functions for the couplings and mass parameters are given by:

$$
\begin{aligned}
(4\pi)^2\beta_{\lambda_\phi} &= 3\lambda_\phi^2 F_2(a_\phi) + 12 g^4 F_2(a_\chi)\,, && (5.71)\\
(4\pi)^2\beta_{\lambda_\chi} &= 3\lambda_\chi^2 F_2(a_\chi) + 12 g^4 F_2(a_\phi) + (8\lambda_\chi h^2 - 48 h^4) N_F\,, && (5.72)\\
(4\pi)^2\beta_{g^2} &= g^2(\lambda_\phi F_2(a_\phi) + \lambda_\chi F_2(a_\chi) + 8 g^2 F_1(a_\chi) + 4 N_F h^2)\,, && (5.73)\\
(4\pi)^2\beta_{h^2} &= h^4(6F_3(a_\chi) + 4N_F)\,, && (5.74)\\
(4\pi)^2\beta_{m_\chi^2} &= 4 N_F h^2 m_\chi^2\,, && (5.75)\\
(4\pi)^2\beta_{m_\phi^2} &= 0\,, && (5.76)\\
(4\pi)^2\beta_{\Omega} &= 0\,, && (5.77)
\end{aligned}
$$

where N_F is the number of massless fermions, and we have defined $a_\alpha = M_\alpha^2/\mu^2$.

Notice that by using a MDR scheme, with no couplings of the field ϕ to fermions, the mass parameter m_ϕ^2 and the vacuum energy contribution Ω do not run: they are fixed by the boundary conditions. That is, the pure (quadratically) divergent terms from vacuum diagrams with no reference to the external energy scale can be subtracted from the bare parameters in the potential, leaving a fixed, finite contribution. For example we can always impose as normalization condition for the vacuum contribution:

$$\Omega(\mu) = 0\,. \tag{5.78}$$

This is equivalent to normalize the effective potential to zero when the field vev vanished. Similarly, the mass parameter $m_\phi(\mu)$ can be set to zero or to the appropriate low mass scale.

We remark that we are studying a model with a light scalar field ϕ, and a much heavier state χ. Our physical initial set-up includes by definition a light mass, whatever the origin, and in particular for inflation or quintessence we know that this has to be at most of order the Hubble scale. This implies that the renormalized mass parameter in a given renormalization scheme must also be of that order, and in particular in the MDS we are using, setting that as an initial condition at a scale μ ensures that at 1-loop order it is fulfilled at any other scale. On the other hand, this can be viewed as a fine-tuning of the bare mass parameter against the quadratic cut-off contribution in order to render the mass parameter of the right order of magnitude. Given that the bare parameters are neither physical, nor related to physical observables, one is always free to choose such initial condition as demanded by the problem one wants to study. In the language of the EFT, usually one appeals to arguments of naturalness to set all the parameters in the model at the high-energy cut-off Λ_0 to be order one in

	c_0	c_1	c_2
$F_1(a)$	15.3	20.1	31.8
$F_2(a)$	48.1	60.04	295.4
$F_3(a)$	48.1	60.04	295.4

units of Λ_0, except for the light state, whose mass has to be fine-tuned [14, 68] below the cut-off. One therefore should check the stability of such an assumption against radiative corrections, or equivalently the renormalization flow. And indeed decoupling ensures that the light states are not destabilized at low-energies by the presence of heavy massive states. The relevant mass parameter in our case is the effective field dependent mass M_ϕ^2, which does run due to the running of the coupling constant λ_ϕ, and the effects of the heavy states on the light mass are encoded in the running of this coupling. During inflation or quintessence in order to have a viable model and a light enough scalar mass this self-coupling has to be tiny, which might also be seen as a fine-tuning problem. Here we are not trying to explain the origin of these 'tiny' parameters when compared to other parameters of the model, but their stability against radiative corrections and the viability of the model in the presence of heavy states.

Coming back to the RGEs Eqs. (5.71-5.75), the expressions for the threshold functions F_i, $i = 1, 2, 3$, are given in Eqs. (A.47) - (A.49) in Appendix A (also some 2-loop expressions are given in Appendix B). They can be well approximated by:

$$F_j(a) \simeq \frac{1 + c_{0j}a}{1 + c_{1j}a + c_{2j}a^2}\,, \tag{5.79}$$

with the coefficients c_{ij} for each function given in Table 5.3. In the massless limit, all threshold functions reduce to one, which recovers for the couplings (dimensionless parameters) the standard RGEs computed for example in $\overline{\text{MS}}$ (see Appendix A). On the other hand, when the ratio $a_\alpha \gg 1$, we have $F_j(a_\alpha) \simeq O(1/a_\alpha)$, i.e., power suppression of the heavy state contribution. Decoupling is not instantaneous, as can be seen in Fig. 5.1. Threshold functions smoothly interpolate between the high energy regime where massive states can be viewed as massless, and the low-energy theory without heavy states.

All that remains now is to fix the initial conditions at some scale μ to integrate the RGEs and obtain the values of parameters when all masses decouple. With those we evaluate the tree-level potential to obtain the RG-improved effective potential at 1-loop. It is shown in Appendix C that by substituting back the solution for the running couplings at low-energy in the tree-level potential one recovers the 1-loop correction

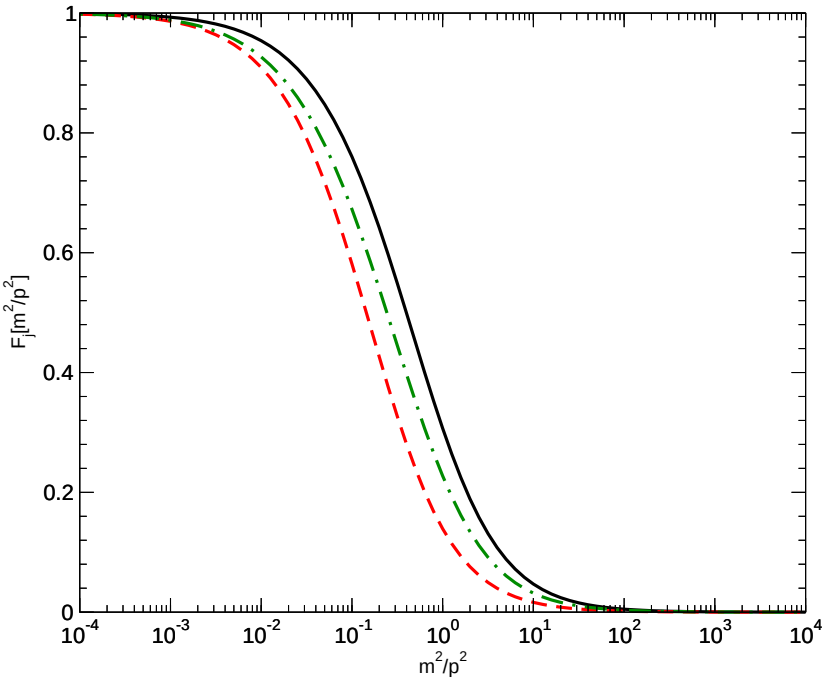

Figure 5.1: Threshold functions: $F_1(m^2/p^2)$ (solid, black), $F_2(m^2/p^2)$ (dashed, red), and $F_3(m^2/p^2)$ (dot-dashed, green).

computed in the MDR scheme:

$$\Delta V^{(1)} = \frac{1}{64\pi^2}\left(\frac{\lambda_\phi^2\phi^4}{4}(\ln\frac{M_\phi^2}{\mu^2} - I(M_\phi^2/\mu^2)) + g^4\phi^4(\ln\frac{M_\chi^2}{\mu^2} - I(M_\chi^2/\mu^2))\right) \quad (5.80)$$

where

$$I(a) = \ln a - 2 - \sqrt{1+4a}\ln\frac{\sqrt{1+4a}-1}{\sqrt{1+4a}+1}\,. \quad (5.81)$$

The main difference between the 1-loop correction computed in a mass independent renormalization procedure, and the MDR scheme one, comes into the non-logarithmic contribution, and it is due to the different scheme-dependent finite contributions in the renormalization conditions. Comparing Eq. (5.80) with Eq. (5.55), the constant "3/2" term is replaced by a threshold function $I(a)$, which controls the contribution of the original log term. Thus, whatever the hierarchy among the masses, we obtain the exact result at 1-loop:

$$\Delta V^{(1)}(\mu = 0) = 0\,. \quad (5.82)$$

In section 5.5 we will present some examples of the procedure for an inflation model, and a quintessence one. We want to check the impact of the radiative corrections on the inflaton/quintessence potential as ϕ changes. Notice that by changing ϕ we are implicitly changing the threshold conditions that depend on the effective masses, and

therefore the values of the couplings, so effectively what we have are field dependent couplings. We will check that at least during the regimes of interest, when ϕ evolves the couplings remain in the perturbative regime and none of them picks up a large correction. For example for the quintessence model, we can impose that we have indeed a quintessence regime, such that for $\phi \gg m_P$ the quartic coupling λ_ϕ is tiny (or even zero). By going backwards in time, i.e., taking smaller values of the field, the evolution should be such that the coupling never gets large enough to disturb the standard evolution of the quintessence field.

5.4 1-loop corrections: non-renormalizable self-interactions

Most quintessence potentials (and some inflation potentials) are typically given by non-renormalizable ϕ potentials, $V_{NR}(\phi)$, so that the tree-level potential is now given by:

$$V^{(0)}(\phi,\chi) = V_{NR}(\phi) + \frac{1}{2}m_\phi^2\phi^2 + \frac{\lambda_\phi}{4!}\phi^4 + \frac{1}{2}m_\chi^2\chi^2 + \frac{\lambda_\chi}{4!}\chi^4 + \frac{1}{2}g^2\chi^2\phi^2\,, \tag{5.83}$$

The ϕ-dependent mass of the ϕ field picks up an extra term due to the non-renormalizable interaction, and in Eq. (5.50) we have to replace M_ϕ^2 by $V''_{NR}+M_\phi^2$. The 1-loop correction $\Delta V^{(1)}$ can be split into a renormalizable and a non-renormalizable contribution, owing to the origin of the field-dependent masses:

$$\Delta V^{(1)} = \Delta V^{(1)}_{NR} + \Delta V^{(1)}_{\rm ren}\,, \tag{5.84}$$

where $\Delta V^{(1)}_{\rm ren}$ is given in Eq. (5.50), with $M_\alpha = M_\phi,\, M_\chi$, while ΔV_{NR} is given by:

$$\begin{aligned}\Delta V^{(1)}_{NR} &= \frac{V''_{NR}}{32\pi^2}\Lambda^2 \\ &+ \frac{1}{64\pi^2}\left(V''_{NR}(V''_{NR}+2M_\phi^2)\left(\ln\frac{V''_{NR}+M_\phi^2}{\Lambda^2}-\frac{1}{2}\right)+M_\phi^4\ln\frac{V''_{NR}+M_\phi^2}{M_\phi^2}\right)\end{aligned} \tag{5.85}$$

Having dealt with the renormalizable interactions in the previous section, we come back to the quadratic cut-off and log dependent term due to the non-renormalizable interaction. The standard approach, which we will follow, is then to consider Λ as the effective ultraviolet cutoff for the model, such that:

$$\Delta V^{(1)}_{NR} \simeq \frac{V''_{NR}}{32\pi^2}\Lambda^2\,, \tag{5.86}$$

where in $\Delta V^{(1)}_{NR}$ we have kept only the dominant quadratic contribution [53].

Without loss of generality, let us consider a generic quintessence potential of the form:

$$V_{NR}\,(\phi) = \frac{\lambda M^{(4+n)}}{\phi^n + M^n}\,. \tag{5.87}$$

The scale M is model dependent, and we do not consider any particular value; here it only parametrizes the value of the field well outside the quintessence phase, $\phi \sim M$, while the quintessence regime happens for values of the field $\phi \gg m_P$. Now we want to check that indeed $\Delta V^{(1)}_{NR}$ does not provide large corrections to the potential, $\Delta V^{(1)}_{NR} \ll V_{NR}$. This condition is not difficult to fulfill when $\phi \gg M$, and in this regime we simply have $\Delta V_{NR}/V_{NR} \simeq (\Lambda/\phi)^2$, so in the quintessence regime when $\phi > m_P$, the mass squared V''_{NR} has become tiny, and the effective ultraviolet cut-off can be taken close to the Planck scale and still $\Lambda/\phi < 1$. However, at early times when $\phi \sim M \ll m_P$ we have that $\Delta V^{(1)}_{NR}/V_{NR} \simeq (\Lambda/M)^2$, which can be large unless the cut-off is well below M.

Nonetheless, one can argue as done in Ref. [26] that this is all right as far as the 1-loop contribution is suppressed not with respect to the tree-level potential, but with respect to the dominant energy density at the time. For example there are some restrictions on the amount of dark energy at the time of Big Bang Nucleosynthesis, with $\rho_{DE} < 0.2\rho_{\rm rad}$, which in this case should be satisfied by the 1-loop effective potential. At earlier times, the quintessence field may find itself fast rolling the potential, with the Universe dominated by its kinetic energy density ρ_{KE} (kination). The condition to be on the safe side would be that still $\Delta V^{(1)}_{NR} \ll \rho_{KE}$. However, it is not clear how to reconcile a fast rolling field with the calculation of the improved effective potential, and the approximation may break down. Because of that we do not pursue the calculation of the effective potential into that regime. Whenever a kination phase due to the quintessence field in the early universe, we can check that after that quantum corrections do not mess up the evolution of the quintessence field, and that quintessence domination is reached today. Going backwards in time, if the quantum corrections are subdominant by the time of kination, we assume that they will not grow as much as to change this phase. We have no means to consistently check this assumption, but we consider it a reasonable working hypothesis.

5.5 Results for inflation and quintessence

5.5.1 Inflation

Let us consider inflationary potentials of the form:

$$V(\phi) = \lambda_\phi M^4 \left(\frac{\phi}{M}\right)^n , \tag{5.88}$$

which is added to $V(\chi)$ in Eq. (5.44). Due to the coupling between ϕ and χ, radiative corrections will always induce a quartic interaction for the inflaton field. Just to keep the discussion simple, we focus on the $n = 4$ case. The tree-level potential reduces then to Eq. (5.45), with[2] $m_\phi = 0$.

In such chaotic potentials inflation takes place for $\phi > m_P$, and therefore the χ field gets a large mass $g\phi \simeq m_P$, for moderate values of the coupling g. On the other hand, the inflaton mass is $M_\phi^2 = V_{\phi\phi} < H^2 \ll M_\chi^2$. Then, following our prescription, the appropriate renormalization scale $\mu = \mu*$ for examining physics during inflation is of the order of M_ϕ. At this scale, the threshold functions imply a suppression of the effect of the χ loops in the renormalization group equations.

Notice that in the standard $\overline{\rm MS}$ scheme, the generic approach is to suppress large logs, in which case at 1-loop order it would mean taking $\mu* \simeq M_\chi$. However, in the MDR scheme the large log contribution is suppressed by the threshold function prefactor when μ is chosen at the lowest threshold. In fact, at 2-loops at order g^4, there will be large logs depending on both masses M_ϕ and M_χ, and in the $\overline{\rm MS}$ scheme it is not clear which is the best choice for $\mu*$. Nevertheless, the MDR scheme is unambiguous that it is around the lowest mass scale.

Up to order of magnitude, the above approach fixes the choice of $\mu*$ in the MDR scheme. But there is an uncertainty in the exact value one should choose. This underlies an inherent ambiguity in the value of the effective potential, which ultimately implies a theoretical uncertainty in the coupling and thus on the model predictions such as the amplitude of primordial perturbations. If the effective potential could be computed exactly, then it would be completely μ-independent and any choice of $\mu*$ would be equally good. However in any real calculation, where the effective potential is calculated only to some finite order, often just 1-loop order, the choice of $\mu*$ must be made carefully. For a given choice of $\mu*$, slightly larger or smaller values should result in the same answer, and if they do not, the selection of $\mu*$ is flawed for the given order in the loop expansion.

To implement the calculation of the RG improved effective potential, two renormalization scales in general are needed. First is the scale, which will be called $\bar{\mu}$, where the initial values of the parameters in the theory are specified and second is the scale where we want to use the potential to do physics, which we have already denoted $\mu*$. A

[2]We have shown in section 5.3 that the mass parameter m_ϕ does not run in the MDR scheme, so that imposing as a boundary condition $m_\phi(\mu) = 0$ ensures that this parameter vanishes at 1-loop at any other energy scale.

further detail in specifying the initial values of the parameters at $\bar{\mu}$ is this specification in general must be given over some range of ϕ, and thus the masses in the system. As ϕ changes, these ϕ-dependent masses will change, thus this choice of $\bar{\mu}$ and/or initial values of the system parameters can change.

There are two approaches we will consider for initializing the RG improved calculation of the effective potential, which will be referred to as the high and low-energy approaches. In the low-energy approach the parameters of the system are initialized at the scale where one is interested in using the effective potential to do physics, thus $\bar{\mu} = \mu*$. Since this $\mu*$ in MDR scheme is at the scale of the smallest mass, it means there would be very small quantum corrections to the effective potential from heavy mass states. Moreover as ϕ changes, thus the masses in the system, in principle one could use RG to move the value of $\mu*$ to optimize the quantum corrections, although the effects would be small. In the high energy approach, the values of the initial parameters are specified at some high renormalization scale $\bar{\mu}$ and then the renormalization group is used to run the parameters down to the scale $\mu*$, where physics is to be done.

An example of implementing the low-energy approach in the case of inflation would be to fix the value of λ_ϕ at the epoch of inflation corresponding today to the largest observable scale. This value of λ_ϕ can be determined from density perturbation constraints from measurement of the cosmic microwave background. The renormalization scale this corresponds to would be (following our prescription) the lowest mass scale in the theory, $\mu* \simeq M_\phi$. We may therefore say that at these values of ϕ and μ, observations tell us $\lambda_\phi(\mu*) \approx 10^{-14}$, and this is all. If we have a more complicated model with additional parameters, we must also be able to specify the remaining parameters such as g^2 also at $\mu*$. In this approach there would be no or very small quantum corrections, so that the tree-level potential would be almost identical to the RG-improved one.

The high energy approach might be implemented if the scalar potential were embedded in a higher theory, and some symmetry at a high energy scale $\bar{\mu}$ specified the value of the parameters and over some range of ϕ. Then the RGE could be used to run the parameters down from $\bar{\mu}$ to $\mu*$ where one wishes to do physics with the effective potential.

An important point here is the matter of initial conditions is not simply a mathematical concern. There is in general missing *physics* in inflationary models. The predictions one obtains from such models depend on this missing physics. Thus for a given inflationary potential, depending what higher theory it is embedded in, different specifications might emerge for the value of the coupling λ_ϕ at some high energy

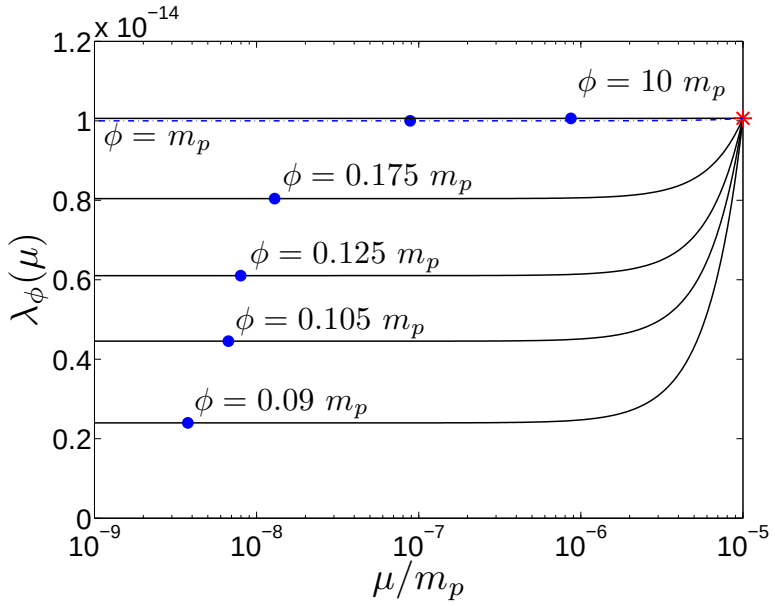

Figure 5.2: The running of λ_ϕ, with $g^2 = 10^{-4}$. The red star indicates $\mu = \bar{\mu}$; at this scale some assumed physics informs us that λ_ϕ is constant over the considered range in ϕ. Blue dots indicate the position on each curve where $\mu = \mu*$, and where the parameter should be taken to improve the effective potential. Curves with smaller ϕ than those shown quickly drive λ_ϕ negative before $\mu*$ is reached. At larger ϕ the curves are flat and constant, with a value that increases only slightly as ϕ increases.

scale, which when evolved down to $\mu*$ will lead to different predictions for inflation and large scale structure. This can be viewed in an alternative fashion. In the low-energy approach, we require the inflation model to be consistent with observation, thus the parameters are determined by the physics of inflation. This might then be used to place constraints on the unspecified physics at higher energies.

Let us now examine the behavior of the parameters with RG running. For the high energy approach at some renormalization scale $\mu = \bar{\mu}$ some unknown physics specifies that the parameters (and λ_ϕ in particular) maintain their values over a range in ϕ. Although we are not considering any specific model for the high scale physics, we would like to investigate the procedure for how this would in principle be done. As such, what we will do to determine the values of the parameters at scale $\bar{\mu}$ is run from $\mu*$ (where the parameters are specified or known) to $\bar{\mu}$, at the value of ϕ where a constraint on the parameters exists. For our purposes this will be $\phi = m_p$. This is just to ensure that the effective potential will match observed constraints at $\phi = m_p$. We would like to check it continues to do so at larger/smaller values of ϕ.

Thus we start with the following parameters at $\phi = m_p$ and $\mu = \mu* = M_\phi$: $\lambda_\phi = 10^{-14}$, $g^2 = 10^{-4}$, $\lambda_\chi = 10^{-3}$, $h^2 = 10^{-4}$, $m_\chi^2 = 10^{-9} m_P^2$, $m_\phi^2 = 0$, $N_F = 8$, and $\bar{\mu} = 10^{-5} m_P$. We run the parameters upwards to $\bar{\mu}$ and find the value of the parameters there. At this scale our assumed physics keeps these parameters unchanged with ϕ. Thus we now change ϕ, and run back down to $\mu = \mu*$. This lets us probe the effective potential at different values of ϕ.

What is found is most parameters remain nearly constant (varying less than 0.1% over the range in μ considered); the only parameter that runs appreciably is λ_ϕ, and is plotted on Fig. 5.2. The blue dashed line on Fig. 5.2 shows this initial curve ($\phi = m_p$) for λ_ϕ. The red star denotes $\bar{\mu}$, which is at the same location on all curves.

Recall at this scale $\bar{\mu}$ the parameters are *assumed* to be independent of ϕ: through this assumption we are including the missing physics. We run down from the red star to the low scale $\mu*$ for different values of ϕ (the blue dots denote the point $\mu*$ on each line). At this scale, the parameter should be taken and inserted into the tree-level potential to generate the 1-loop improved effective potential.

It is clear the curves are very flat at the scale $\mu*$; it makes little difference if the parameters are evaluated at precisely $\mu*$ or within some order of magnitude of this scale (or indeed, many orders of magnitude below it). This is precisely as we would expect, as we have chosen $\mu*$ just for this property, so as to satisfy the μ-independence of our improved effective potential.

As each curve generates a single value of $\lambda(\mu*)$ for a single value of ϕ, we can construct curves of how $\lambda(\mu*)$ varies over ϕ directly. To do so we compare the values at $\mu*$ (the blue dots) for a large number of curves. Furthermore, we can examine how these curves change if the value of $\bar{\mu}$ (the horizontal position of the red star) is changed. The result is Fig. 5.3, maintaining the above parameters but varying $\bar{\mu}$.

How can we understand this behavior? Consider the equation governing the running of λ_ϕ, Eq. (5.71),

$$(4\pi)^2 \beta_{\lambda_\phi} = 3\lambda_\phi^2 F_2(a_\phi) + 12 g^4 F_2(a_\chi)\,. \tag{5.89}$$

We will now make some approximations. If we assume $\mu^2 \ll M_\chi^2$ (certainly true for the numerical range shown above), we may use the approximate behavior of the threshold function, $F_2(a_\chi) \approx \mu^2/(g^2\phi^2)$. Despite this suppression, the value of λ_ϕ is still small enough for the second driving term to dominate. Indeed, for such small values of the self-coupling $\lambda_\phi(\mu) \simeq 10^{-14}$, in the absence of any other coupling the solution to the RGE gives $\lambda_\phi(\mu) \simeq \lambda_\phi(\bar{\mu})$, so that effectively $\beta_{\lambda_\phi} \simeq 0$, and the value of the coupling must be initialized at that small value at any other high energy scale. The initial value

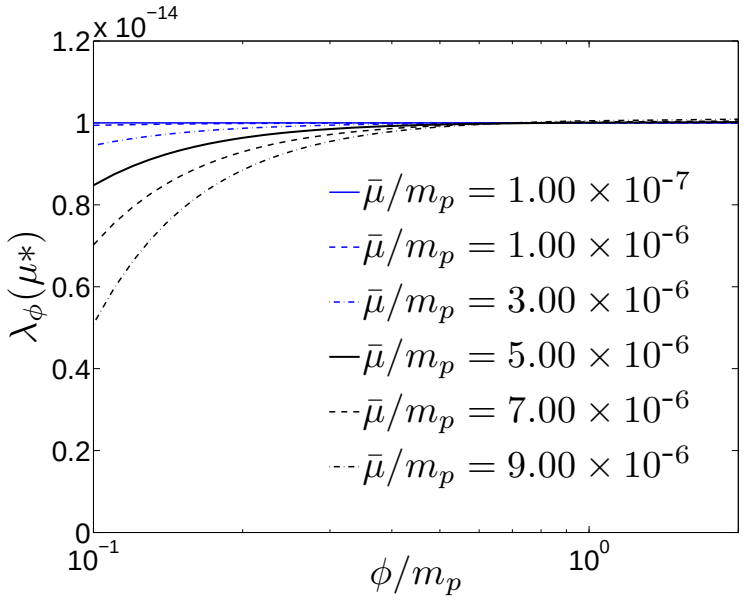

Figure 5.3: Curves showing how $\lambda_\phi(\mu*)$ varies with ϕ, for different values of $\bar{\mu}$. As before, $g^2 = 10^{-4}$. Since $\lambda_\phi(\mu*)$ appears in the RG-improved effective potential, this is the physically important quantity. Curves with $\bar{\mu}$ larger than those shown quickly drive $\lambda_\phi(\mu*)$ negative before ϕ is as small as $0.1m_p$, and have mildly larger values at large ϕ. At larger ϕ the curves asymptote to a constant value. Increasing g^2 has a similar effect to increasing $\bar{\mu}^2$, and the increase in one may be compensated for by a decrease in the other, allowing for larger values of g^2 than plotted here.

of λ_ϕ must be fixed with the same accuracy than the given small value at low energies. On the other hand, including the effect of g we find:

$$(4\pi)^2 \frac{d\lambda_\phi}{d\mu} \simeq 12 \frac{g^2}{\phi^2} \mu \,. \tag{5.90}$$

We can solve this if we make the assumption g^2 is constant, which is well supported numerically. The solution is

$$\lambda_\phi(\mu) \simeq \lambda_\phi(\bar{\mu}) - \frac{6g^2}{(4\pi)^2} \left(\frac{\bar{\mu}}{\phi}\right)^2 \left[1 - \left(\frac{\mu}{\bar{\mu}}\right)^2\right] . \tag{5.91}$$

We see now precisely the behavior in Fig. (5.2); starting from an initial value at $\lambda_\phi(\bar{\mu})$, λ_ϕ quickly reaches a constant value when $\mu \ll \bar{\mu}$. We also see the ϕ dependence exhibited in Fig. (5.3). As ϕ is steadily reduced, the second term becomes larger, eventually dominating the first term and driving λ_ϕ negative. Larger values of $\bar{\mu}$ have the same effect as this. For $\lambda_\phi(\mu*)$ to be stable against changes in ϕ, we require the second term smaller than the first. This gives the constraint:

$$\frac{6g^2}{(4\pi)^2} \left(\frac{\bar{\mu}}{\phi}\right)^2 \ll \lambda_\phi(\bar{\mu}) \,, \tag{5.92}$$

which can be written as:

$$g\bar{\mu} \gg 10 M_\phi(\bar{\mu}) \,. \tag{5.93}$$

For inflation, with $\lambda_\phi(\bar{\mu}) \approx 10^{-14}$ and at its smallest $\phi = 0.1 m_p$, we find:

$$g\bar{\mu} \ll 10^{12}\,\mathrm{GeV} \,. \tag{5.94}$$

This agrees well with Fig. 5.3; if $g^2 \approx 10^{-4}$, then $\bar{\mu}$ may be no larger than 10^{14} GeV.

When inflation ends at around $\phi \sim m_p$, the suppression is reduced compared to earlier, larger field values. As the field strength continues to decrease, the effects become gradually larger. The values of $\bar{\mu}$ and g^2 in the plots have been chosen to display this behavior between $\phi = m_p$ and $\phi = 0.1 m_p$.

For completeness, note that if a large mass parameter is given to χ, such that $M_\chi \approx m_\chi \gg g^2\phi^2$, the solution becomes independent of ϕ at these field values. However, for the solution to be valid, we would require $M_\chi \approx m_\chi \gg \bar{\mu}$ to suppress the χ loops.

For a given value of g, Eq. (5.93) tell us which is the effective high energy scale for the model below which the coupling g does not induce large corrections to the inflaton self-coupling, i. e., the effects of the heavy field decouple. Similarly, choosing the value of $\bar{\mu}$, it tells us how large can be the value of g. On the other hand, it is just an initial

condition problem to find the value of $g(\bar{\mu})$ (having also set the values of the other couplings) which gives rise to the required value of $\lambda_\phi(\mu)$. The problem is how much that initial value has to be fine-tuned in order not to induce a variation on $\lambda_\phi(\mu)$ say larger than 10 %, a question that can be answered with the RGEs. Taking for example $\phi = m_P$, for the GUT scale with $\bar{\mu} \simeq 10^{16}$ GeV, an initial value $g^2(\bar{\mu}) \simeq 10^{-6}$ does not require any tuning, while a value $g^2(\bar{\mu}) \simeq 10^{-4}$ requires to be fixed with an accuracy of less than 10^{-4}%. With $\bar{\mu} = m_P$, we obtain that $g^2(\bar{\mu}) \simeq 10^{-7}$ does not require any tuning, but $g^2(\bar{\mu}) \simeq 10^{-4}$ has to be fixed with an accuracy of less than 10^{-6}%. Without taking into account thresholds in the RGEs, either with $\bar{\mu}$ the GUT scale or the Planck scale, one would conclude that $g^2(\bar{\mu})$ has to be fine-tuned within a 10^{-7}% of its initial value.

Let us now pause and consider the physical interpretation of these results. Without decoupling, we would expect corrections to λ_ϕ of size $\mathcal{O}(g^4)$. Instead, due to the suppression of the χ-loops, we find this result is reduced by a factor of $\bar{\mu}^2/M_\chi^2$. Deriving an effective field theory, by integrating out the heavy degrees of freedom from a more fundamental theory, typically leaves behind a low-energy theory along with terms in the potential suppressed by factors of $\mathcal{O}\left(E^2/m^2\right)$, where E is the low-energy scale the theory is probed at, while m is the energy scale associated with the heavy degrees of freedom. For instance, classical electromagnetism encounters corrections with a size $\mathcal{O}\left(E^2/m_e^2\right)$ where m_e is the mass of the electron. We might have expected in the case of inflation to find quantum corrections suppressed in a similar fashion, with corrections $\mathcal{O}\left(E^2/M_\chi^2\right)$. But when dealing with an effective potential directly, it is not clear what energy scale to associate to E. The effective potential is computed from diagrams with external legs set to zero momentum, and the only other scale is the renormalization scale, upon which our results cannot depend. We now have an answer (at least for the form we have assumed our missing physics take): it is the scale associated with $\bar{\mu}$, where the theory is set. If the extremely large field-strength associated with the inflaton (with $\phi > m_p$ in chaotic inflation) serve to generate field dependent mass of the χ-field far above this scale, its contribution to the effective potential is suppressed.

At what scale should we expect $\bar{\mu}$ to be? Inflation is associated with high vacuum energy densities. But the relationship between the renormalization scale and physical energy scale is not clear. Without an obvious way of relating these two properties, it is difficult to motivate any particular choice of $\bar{\mu}$ above any other.

5.5.2 Quintessence

Now we revert to the generic quintessence potential introduced in Eq.(5.87). Quintessence occurs when $\phi \gg m_P$, and thus, any field at least moderately coupled to ϕ will acquire a large mass from the ϕ background field value. Therefore, again one expects only power corrections to the effective potential from such heavy states. From the previous example, we can perhaps already guess that we should expect the typical g^4 corrections to λ_ϕ to be suppressed by $\mathcal{O}(\bar{\mu}^2/M_\chi^2)$.

In a similar fashion to inflation, we have only limited knowledge of the effective potential. We know only that today, with $\phi \simeq m_p$, the value of λ_ϕ appearing in the effective potential must be small enough so that the dominant contribution comes from the non-renormalizable term $V_{NR}(\phi)$. Another way of stating this is that the tiny effective field mass of $M_\phi \sim 10^{-33}$ eV remains unchanged by the size of $\lambda_\phi(\mu*)$. From this, we can place a constraint on $\lambda_\phi(\mu*) \ll 10^{-124}$.

At earlier epochs in the history of the universe, the quintessence field strength is smaller. As discussed earlier, at very small field strengths the effective potential description breaks down, but we would like to make sure the potential is not disrupted by quantum corrections over at least an order of magnitude in ϕ.

We proceed in just the same way as in the previous section with inflation. The system of equations are initialized at $\mu = \mu*$ with $\phi = m_p$. We take parameters at this scale as: $\lambda_\phi = 10^{-126}$, $g^2 = 10^{-20}$, $\lambda_\chi = 10^{-3}$, $h^2 = 10^{-4}$, $m_\chi^2 = 10^{-9} m_p^2$, $m_\phi^2 = 0$, $N_F = 8$, and $\bar{\mu} = 10^{-57} m_p$. All is identical as with inflation, except for the size of λ_ϕ, g^2 and $\bar{\mu}$, and of course the non-renormalizable contribution to M_ϕ. The result is Fig. 5.4, where once again the blue dashed line indicates the initial curve ($\phi = m_p$) that we use to find appropriate parameters at $\bar{\mu}$, so that (by construction) we match observational constraints at this value of ϕ. The red star indicates $\bar{\mu}$ where the initial curve ends, and all other curves (each for a different value of ϕ) begin. Blue dots indicate the location of $\mu*$ on each curve.

The results are very similar to inflation. Smaller values of ϕ cause $\lambda_\phi(\mu*)$ to rapidly approach zero. Due to M_ϕ being proportional to an inverse power of ϕ, decreasing ϕ also increases $\mu* = M_\phi$. Thus the blue dots on each curve shift to the right as ϕ is decreased, the opposite for inflation.

Making ϕ small enough eventually drives $\lambda_\phi(\mu*)$ past zero. Our concern for quintessence is different to that of inflation; we are not concerned with λ_ϕ becoming negative; rather we are concerned with it becoming *large* compared with the non-

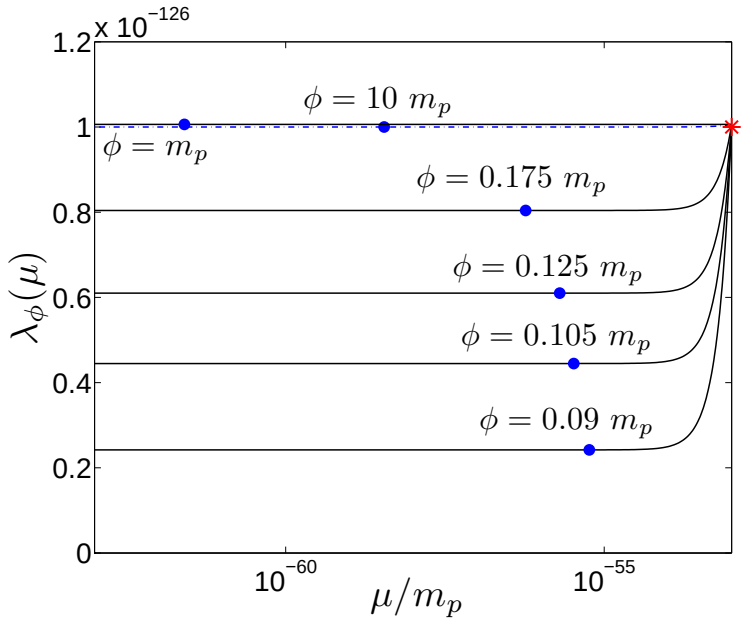

Figure 5.4: The running of λ_ϕ, with $g^2 = 10^{-20}$. As with inflation, the red star indicates $\bar{\mu}$ where the parameters are assumed to be ϕ independent. The blue dots on each curve indicate $\mu*$ for each value of ϕ. Increasing g^2 has a similar effect to decreasing ϕ^2.

renormalizable term in the potential. However, once $\lambda_\phi(\mu*)$ becomes negative, smaller values of ϕ decrease it further, such that its absolute value becomes large enough to be problematic. Thus, to ensure the viability of the quintessence model, it is sufficient to make the same demand as with inflation: that λ_ϕ remains positive for a considered range in ϕ.

Just as before, we construct curves of $\lambda_\phi(\mu*)$ with respect to ϕ, shown on Fig. 5.4. As with inflation, very small values of ϕ likely signal a breakdown in the description of the field in terms of an effective potential. This is particularly clear in the case of quintessence, where in the early universe and at small field values $\phi \ll m_p$ the field may be moving extremely quickly, to the extent that the field's behavior is dominated by the kinetic terms. Again at larger field values, there is little difference to the value of the parameter. Due to the non-renormalizable term in the potential, smaller values of ϕ actually leads to a larger value of $\mu*$. This makes little difference to the behavior of $\lambda_\phi(\mu*)$ with respect to ϕ however, as the curves are so flat at low values of μ. Fig. 5.5 shows this dependence of $\lambda_\phi(\mu*)$ with ϕ, for different values of $\bar{\mu}$. The behavior is much the same, and can be understood in precisely the same way.

The equations remain unchanged when moving from inflation to quintessence. The solution is therefore identical, and following the same line of reasoning, we may therefore

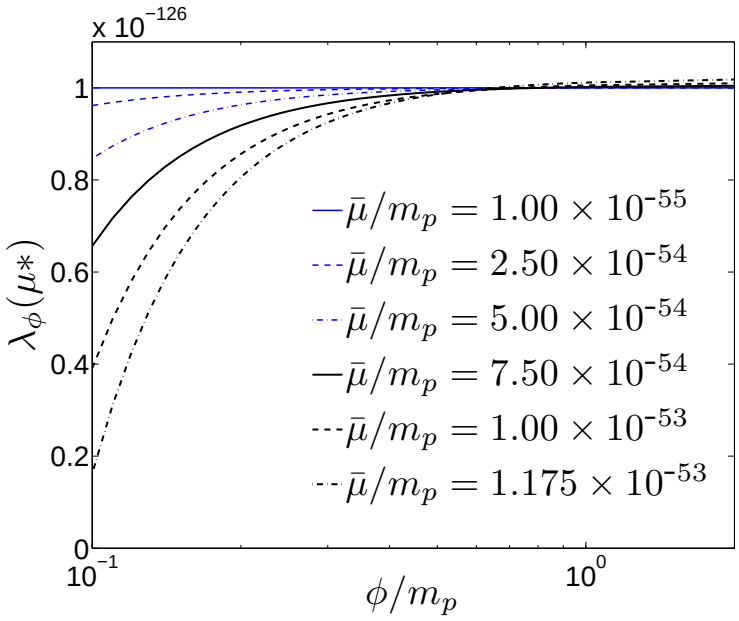

Figure 5.5: Curves of $\lambda_\phi(\mu*)$ against $\bar{\mu}$. The value of g^2 remains at 10^{-20}. At large values of ϕ the curves become flat, while small values induce large changes in $\lambda_\phi(\mu*)$, eventually driving it negative (and eventually, large). As with inflation, larger values of g^2 may be compensated for by using smaller values of $\bar{\mu}$, and vice-versa.

write down the constraint:

$$g\bar{\mu} \ll 10 M_\phi(\bar{\mu}) \simeq 10^{-42}\,\mathrm{GeV}\,. \tag{5.95}$$

Exactly how constrained the coupling g^2 is, depends upon the renormalization scale $\bar{\mu}$ at which the physics is set. Keeping the latter within the range of the quintessence mass, one can work safely with values of the coupling $g^2 \simeq 10^{-4}$. Forcing the high energy scale to be too many orders of magnitude larger than the quintessence mass will require the coupling g^2 to be extremely suppressed.

Quintessence is a low-energy phenomena, operating at scales great many orders of magnitude lower than inflation. The constraint on λ_ϕ means that the combination $g\bar{\mu}$ must be much smaller than in the case of inflation, to avoid generating corrections to the effective potential that ruins the tree-level behavior. However once again there is a great deal of ambiguity as to what an appropriate value of $\bar{\mu}$ should be. This is information not included in quintessence models, just as it is not included in models of inflation.

5.6 Summary

The approach in this chapter is consistent with the concepts of low-energy effective field theories [119, 13, 14, 68, 29, 120, 107]. In the effective field theory approach, at low energies, reflected by a low renormalization scale, the effective theory is obtained from the full theory by removing all heavy mass field internal propagator lines and treating the heavy fields only as external particles in Green's functions, with all other effects emerging through renomalization of the parameters in the theory. In a scalar field model, decoupling of heavy states from the low energy effective action given by the light degrees of freedom was probed in [14, 68]. In this chapter, these concepts have been applied to the effective potential, in application to inflation and quintessence models. This follows previous treatments of decoupling in effective potentials, applied to electroweak physics [16, 112, 34]. What is proposed here is that in interacting inflation or quintessence models, if there are disparate mass scales in the theory, then a mass independent renormalization scheme, such as the MS or $\overline{\rm MS}$ schemes is not adequate in capturing the physics of decoupling. Rather the mass dependent renormalization scheme is more appropriate. This approach has already been applied to other problems in physics, like for example the Higgs sector of the Minimal Supersymmetric Standard Model (MSSM) [51, 110]. By using a mass-dependent subtraction scheme, the on-shell renormalization scheme [48], it was shown that in the so-called decoupling limit, when all the scalar masses other than the lightest Higgs are heavier than the electroweak scale, one effectively recovers the Higgs sector of the Standard Model at the electroweak scale. That is, the effects of new physics decouple at the electroweak scale. Similarly, in the mass dependent scheme used in this chapter, we have shown that heavy fields coupled to the inflaton or quintessence field have their quantum effects power suppressed in the effective potential.

To implement the MDR scheme, the choice of renormalization scale, and thus the division between low and high energy scales in the effective potential, must be determined. Although the effective potential must be independent of the renormalization scale, the issue is at what scale should the parameters in the theory be initialized. As noted in this chapter, this is an underlying ambiguity in inflation and quintessence models and can lead to differing results. We discussed two options which we called the low-energy and high-energy approaches in Sect. 5.5. In the low-energy approach the parameters are initialized at a renormalization scale of the order of the lowest mass scale in the system. One consequence of this approach is, if the inflaton or quintessence

field is the field with lowest mass, and all other fields interacting with it are much heavier, then there will be negligible quantum corrections, due to the decoupling effects. In the EFT description, the low-energy effective potential is given only by the light degrees of freedom. Alternative in the high-energy approach, the parameters are initialized at some high energy scale, possibly if the theory were embedded in some higher theory and some symmetry principle determined the parameters at high scale. In this case, one can use the renormalization group equations to evolve the scale to where one wants to do physics. By demanding the radiative corrections due to the heavy fields not to spoil the flatness of the potential, i.e., keeping the self-coupling λ_ϕ small enough, one can derive the energy scale range of validity of the effective theory, as given in Eqs. (5.93) and (5.95). Due to decoupling, this constraint on the coupling of the heavy to the light state is milder than the constraint one would derive without decoupling, and it is directly related to the energy scale upper limit for the effective field theory. Nevertheless, in the absence of the inflation or quintessence model being embedded in some higher theory, if the model has disparate mass scale, the low-energy approach is technically a viable option, in which case even if the model is moderately interacting with other fields, because of the decoupling effect it can still produce the very flat potentials needed.

In cosmology there has been a common practice in the treatment of the scalar field effective potential, whereby the effect at 1-loop order of any quantum field coupled to this scalar field leads to a Coleman-Weinberg correction term in the quantum corrected effective potential [102]. However this procedure is not always applicable. In particular, when the mass of a quantum field is greatly in excess of all other masses in the system, the decoupling theorem implies the quantum corrections from this heavy field are suppressed in powers of the light-to-heavy mass ratio.

There are many common models in inflation and quintessence where these suppression effects become valid. For inflation, in all large field models, often terms chaotic inflation models [96], as well as some hybrid models [98] inflation occurs when the inflaton background field value is very large $\langle\phi\rangle \sim m_p$. Thus any scalar or fermion field coupled to the inflaton with at least moderate coupling will have a very large mass, much bigger than the inflaton mass. The quantum correction from such fields will thus be greatly suppressed during the inflation period due to decoupling. Also in small field models [3, 97] after inflation, as the inflaton background field value grows, it is possible fields coupled to the inflaton acquire large mass and thus there quantum corrections might become suppressed. In quintessence models, the dark energy regime

in most models occurs when the scalar field background value is very large $\phi \sim m_p$ [114, 152, 30, 31, 27, 138, 20, 71, 115]. Thus once again during the quintessence regime, quantum corrections from other scalar and fermion fields coupled to the quintessence field will be highly suppressed from decoupling effects.

This decoupling effect can have significant importance to the building of inflation and quintessence models. For inflation it implies the inflaton field can be coupled more strongly to other heavy fields. Stronger couplings can have beneficial effects in leading to more robust reheating after inflation or in the case of warm inflation models [19], can lead to a greater parameter regime and many more possible models. In quintessence models, the potential during the dark energy regime typically has to be so flat, that the quintessence field often is just added as an additional field without any other dynamical purpose aside from driving the late time dark energy expansion of the universe. However the decoupling effect means that at early times the quintessence field might be interacting with other fields and produce dynamical consequences and at later time as the quintessence field background value becomes large, these other fields become very massive, thus decoupling, which then leads to an almost noninteracting field with the required ultra-flat potential to drive the late time dark energy phase.

Chapter 6

Cosmological perturbation theory

6.1 Introduction

The universe is not quite homogeneous, but so far we have not looked at all at how these inhomogeneities evolve with time. We would expect overdensities to grow with time as they attract more matter, and underdensities to shrink, but the expansion of the universe will obviously play some role in slowing the growth of structure.

Trying to describe the universe in terms of the exact pattern of overdensities we observe on the sky is obviously unreasonable, but an excellent simplification is to describe the characteristic fluctuations in term of their Fourier transform. In this sense the statistical behaviour of fluctuations can be examined, without worrying about the complicated exact deviations from homogeneity that we observe.

Another simplification is to only consider the behaviour at a linear level, in the sense that when expanding out quantities such as the density ρ as a homogeneous part $\bar{\rho}$ and a perturbation $\delta\rho$, only terms linear in the perturbation are kept. Quadratic terms and higher are discarded as small. Provided $\delta \equiv \delta_\rho/\rho \ll 1$, this should be an excellent approximation. When described in terms of Fourier modes k, this means that each k mode behaves independently of every other.

6.2 Perturbed metric

Perturbing the matter distribution will introduce perturbations in the gravitational fields (the components of the metric tensor). There is an inherent ambiguity in describing these perturbations, stemming from the fact that once the universe is no longer perfectly homogeneous, there is no longer a true 'cosmic' time any more. This inherent freedom in the description of the physics is referred to as gauge freedom.

There are three sorts of degrees of freedom to the perturbed FLRW metric. These are tensor, vector and scalar modes. The former two are usually neglected as they usually decay away quickly. We will only consider the scalar modes also.

For our choice of gauge, we use the Newtonian gauge. The two scalar degrees of freedom are Φ and Ψ, and appear in the metric as

$$-ds^2 = dt^2(1+2\Psi) - a^2(1-2\Phi)\gamma_{ij}dx^i dx^j. \tag{6.1}$$

Here γ_{ij} is the metric tensor for flat, Euclidean space.

6.3 Christoffel symbols

The next step is to calculate the Christoffel symbols associated with the perturbed metric. By doing so, we can evaluate the expression $\nabla_\mu T^{\mu\nu} = 0$, the conservation of the energy-momentum tensor for the system.

Dots denote derivatives with respect to coordinate time, primes with respect to conformal time τ ($a\mathrm{d}\tau \equiv \mathrm{d}t$) , and as mentioned earlier we will work in Fourier space so that $\partial_i \to ik_i/a = ik^i/a$, and $\partial_i\partial^i = -k^2/a^2$.

Christoffel symbols are defined by the equation:

$$\Gamma^\gamma_{\mu\nu} = \frac{1}{2}g^{\eta\gamma}\left(\partial_\mu g_{\eta\nu} + \partial_\nu g_{\eta\mu} - \partial_\eta g_{\mu\nu}\right). \tag{6.2}$$

First consider $\gamma = 0$. This gives,

$$\Gamma^0_{\mu\nu} = \frac{1}{2}g^{\eta 0}\left(\partial_\mu g_{\eta\nu} + \partial_\nu g_{\eta\mu} - \partial_\eta g_{\mu\nu}\right).$$

Now examine $\mu = \nu = 0$. All the terms are identical, and so

$$\Gamma^0_{00} = \frac{1}{2}(1-2\Psi)2\dot{\Psi} = \dot{\Psi},$$

to first order. Now consider the case where $\mu = 0$ or $\nu = 0$, while the other is equal to $i = 1, 2, 3$. (It does not matter which, as the symbols are symmetric about their lower indices).

$$\Gamma^0_{0i} = \frac{1}{2}(1-2\Psi)\left(\partial_0 g_{0i} + \partial_i g_{00} - \partial_0 g_{0i}\right) = \partial_i\Psi = ik_i\Psi.$$

With both lower indices non-zero,

$$\begin{aligned}\Gamma^0_{ij} =& \frac{1}{2}(1-2_\Psi)\left(\partial_i g_{0j} + \partial_j g_{i0} - \partial_0 g_{ij}\right)\\ =& -\delta_{ij}\left(a^2\dot{\Phi} - a^2 H(1-2\Phi)\right)(1-2\Psi)\\ =& \delta_{ij}\left(H - 2H(\Phi+\Psi) - \dot{\Phi}\right)\end{aligned}$$

The remaining symbols are $\Gamma^i_{\mu,\nu}$. Consider $\mu = \nu = 0$.

$$\begin{aligned}\Gamma^i_{00} =& \frac{1}{2}g^{i\eta}\left(\partial_0 g_{0\eta} + \partial_0 g_{\eta 0} - \partial_\eta g_{00}\right)\\ =& -g^{ij}\partial_j\Psi = \frac{i\Psi k^i}{a^2}\end{aligned}$$

And again, if either $\mu = i$ or $\nu = i$,

$$\begin{aligned}\Gamma^i_{j0} =& \frac{1}{2}g^{i\eta}\left(\partial_j g_{\eta 0} + \partial_0 g_{j\eta} - \partial_\eta g_{j0}\right)\\ =& \frac{1}{2}g^{ik}\partial_0 g_{jk} = \delta^i_j\left(H - \dot{\Phi}\right)\end{aligned}$$

The last symbol is somewhat longer:

$$\Gamma^i_{jk} = \frac{1}{2}g^{i\eta}\left(\partial_j g_{\eta k} + \partial_k g_{\eta j} - \partial_\eta g_{jk}\right) = \frac{1}{2}g^{im}\left(\partial_j g_{mk} + \partial_k g_{mj} - \partial_m g_{jk}\right)$$

Proceeding term by term,

$$\begin{aligned}\frac{1}{2}g^{im}\partial_j g_{mk} =& \delta_{mk}\delta^{im}\frac{1}{2}a^{-2}\partial_j\left(-(1-2\Phi)a^2\right)\\ =& -\Phi i\delta^i_k k_j,\end{aligned}$$

$$\begin{aligned}\frac{1}{2}g^{im}\partial_j g_{mj} &= -\Phi i\delta^i_j k_k,\\ -\frac{1}{2}g^{i\eta}\partial_m g_{jk} &= i\Phi\delta_{jk}k_i\Phi\end{aligned}$$

Putting these together again,

$$\Gamma^i_{jk} = -i\Phi\left(\delta^i_k k_j + \delta^i_j k_k - \delta_{jk}k_i\right).$$

The Christoffel symbols are now also presented here together, for convenient reference.

$$\Gamma^0_{00} = \dot{\Psi} \tag{6.3}$$

$$\Gamma^0_{0i} = ik_i\Psi \tag{6.4}$$

$$\Gamma^0_{ij} = \delta_{ij}a^2\left(H - 2H\left(\Phi+\Psi\right) - \dot{\Phi}\right) \tag{6.5}$$

$$\Gamma^i_{00} = i\Psi\frac{k^i}{a^2} \tag{6.6}$$

$$\Gamma^i_{j0} = \delta^i_j\left(H - \dot{\Phi}\right) \tag{6.7}$$

$$\Gamma^i_{jk} = -i\Phi\left(\delta^i_k k_j + \delta^i_j k_k - \delta_{jk}k^i\right) \tag{6.8}$$

6.4 Conservation of the perturbed energy-momentum tensor

The energy-momentum tensor is given by:

$$T^\mu{}_\nu = (\rho + P)\,U^\mu U_\nu - \delta^\mu_\nu P, \tag{6.9}$$

where U^μ is the four velocity of the fluid. In the rest frame of the fluid, $U^i = 0$ (by definition), and so normalisation $U^\mu U_\mu = 1$ requires $U^0 = (1 - \Psi)$. This is also true to first order if U^i are treated as small.

The coordinate 3-velocity is defined as

$$v^i \equiv \frac{adx^i}{dt} = \frac{U^i}{aU^0} \tag{6.10}$$

with an index that can be raised and lowered using γ_{ij}. Inverting the equation

$$U^i = \frac{v^i}{a}, \quad U_i = -av_i. \tag{6.11}$$

The vector v^i is not *exactly* the proper 3-velocity because adx^i is not proper distance, and dt not proper time, but the corrections are only first order in the metric perturbations.

Conservation of the energy-momentum tensor requires

$$\nabla_\mu T^\mu{}_\nu = 0, \tag{6.12}$$

or for a particular fluid

$$\nabla_\mu T^\mu{}_\nu = Q_\nu, \tag{6.13}$$

where Q_ν is a four-vector describing energy-momentum exchange with other fluids in the system. The sum of all Q_ν for all fluids must naturally be zero, to conserve energy-momentum across the whole system (we will return to this point).

We will be quite general, and allow Q_ν to be non-zero.

Expanding out,

$$\nabla_\mu T^\mu{}_\nu = \partial_\mu T^\mu{}_\nu + \Gamma^\mu_{\eta\mu}T^\eta{}_\nu - \Gamma^\eta_{\nu\mu}T^\mu{}_\eta.$$

6.4.1 δ-equation

Consider the $\nu = 0$ case. Looking at each term in turn, we begin with the first term:

$$\partial_\mu T^\mu{}_0 = \partial_0 T^0{}_0 + \partial_i T^i{}_0 = \dot{\rho} + \partial_i \frac{v^i}{a} (\rho + P) ,$$

Or with derivatives with respect to conformal time denoted by a prime, and $\partial_i v^i = \theta$,

$$\partial_\mu T^\mu{}_0 = \frac{\rho'}{a} + \frac{\theta}{a} (\rho + P) . \tag{6.14}$$

The second term can be broken up,

$$\Gamma^\mu_{\eta\mu} T^\eta{}_0 = \Gamma^\mu_{0\mu} T^0{}_0 + \Gamma^\mu_{i\mu} T^i{}_0.$$

and each term can be further separated,

$$\Gamma^\mu_{\eta\mu} T^\eta{}_0 = \Gamma^0_{00} T^0{}_0 + \Gamma^i_{0i} T^0{}_0 + \Gamma^0_{i0} T^i{}_0 + \Gamma^j_{ij} T^i{}_0.$$

Discarding the last two terms as second order, this yields

$$\Gamma^\mu_{\eta\mu} T^\eta{}_0 = \Psi' \frac{\rho}{a} + 3 \left(\mathcal{H} - \Phi' \right) \frac{\rho}{a} \tag{6.15}$$

The third term:

$$-\Gamma^\eta_{0\mu} T^\mu{}_\mu = -\Gamma^0_{00} T^0{}_0 - \Gamma^0_{0i} T^i{}_0 - \Gamma^i_{00} T^0{}_i - \Gamma^i_{0j} T^j{}_i.$$

The middle two terms are second order, while the remaining two give

$$-\Gamma^\eta_{0\mu} T^\mu{}_\mu = -\Psi' \frac{\rho}{a} + 3 \left(\mathcal{H} - \Phi' \right) \frac{P}{a}.$$

Combining the equations together gives

$$\rho' + \theta (\rho + P) + 3\mathcal{H}(\rho + P) - 3\Phi' (\rho + P) = a Q_0. \tag{6.16}$$

From this, the equation governing the homogeneous background can be extracted:

$$\bar{\rho}' + 3\mathcal{H}(\bar{\rho} + \bar{P}) = a \bar{Q}_0. \tag{6.17}$$

Writing $\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}$, and subtracting the background equation,

$$\delta' + \delta \frac{\bar{\rho}'}{\bar{\rho}} + \theta(1 + w) + 3\mathcal{H} \left(\frac{\delta \bar{P}}{\delta \bar{\rho}} + 1 \right) \delta - 3\Phi' (1 + w) = \frac{a \delta Q_0}{\bar{\rho}},$$

and using $\frac{\bar{\rho}'}{\bar{\rho}} = \frac{a \bar{Q}_0}{\bar{\rho}} - 3\mathcal{H}(1 + w)$, the final result is:

$$\delta' + (1 + w) \left(\theta - 3\Phi' \right) + 3\mathcal{H} \left(\frac{\delta \bar{P}}{\delta \bar{\rho}} - w \right) \delta = -\delta \frac{a \bar{Q}_0}{\bar{\rho}} + \frac{a \delta Q_0}{\bar{\rho}} \tag{6.18}$$

6.4.2 θ-equation

Now consider $\nu = i$. Again, term by term, the first term reads:

$$\partial_\mu T^\mu{}_i = \partial_0 T^0{}_i + \partial_j T^j{}_i = -\left(\dot{\rho} + \dot{P}\right) a v_i - (\rho + P)\left[a\dot{v}_i + aHv_i + \partial_i P\right].$$

Taking the partial (not covariant) derivative of this equation, and recalling $\partial_i v^i = \theta$, the result in conformal time reads:

$$\partial^i \partial_\mu T^\mu{}_i = \left(\dot{\rho} + \dot{P}\right)\frac{\theta}{a} + (\rho + P)\left[\frac{\theta'}{a^2} + \mathcal{H}\frac{\theta}{a^2} - \frac{k^2}{a^2}P\right]. \tag{6.19}$$

The second term:

$$\Gamma^\mu_{\eta\mu} T^\eta{}_i = \Gamma^0_{\eta 0} T^\eta{}_i + \Gamma^j_{\eta j} T^\eta{}_i,$$

which can be again be separated yet further:

$$\Gamma^\mu_{\eta\mu} T^\eta{}_i = \Gamma^0_{00} T^0_i + \Gamma^0_{j0} T^j{}_i + \Gamma^j_{0j} T^0{}_i + \Gamma^j_{kj} T^k{}_i.$$

The first term is second order. After a little algebra, and keeping only those terms first order in the perturbations,

$$\Gamma^\mu_{\eta\mu} T^\eta{}_i = -i\Psi k_i P - 3H(\rho + P) a v_i + i3\Phi k_i P,$$

so that once again taking the partial derivative,

$$\partial^i \Gamma^\mu_{\eta\mu} T^\eta{}_i = -\frac{\Psi k^2}{a^2}P + 3\mathcal{H}(\rho + P)\frac{\theta}{a^2} + 3\frac{\Phi k^2}{a^2}P. \tag{6.20}$$

The third and final term:

$$-\Gamma^\eta_{i\mu} T^\mu{}_\eta = -\Gamma^0_{i0} T^0{}_0 - \Gamma^j_{i0} T^0{}_j - \Gamma^k_{ij} T^j{}_k - \Gamma^0_{ij} T^j{}_0.$$

After a little algebra, and again taking a partial derivative:

$$-\partial^i \Gamma^\eta_{i\mu} T^\mu{}_\eta = -\rho\frac{\Psi k^2}{a^2} - 3\frac{\Phi k^2}{a^2}P \tag{6.21}$$

Using the continuity equation to replace $\bar{\rho}'$, dividing through by $(1 + w)\bar{\rho}$, and re-arranging, the final equation reads:

$$\theta' + \mathcal{H}(1 - 3w)\theta + \frac{w'}{1 + w}\theta - \frac{\delta\bar{P}/\delta\bar{\rho}}{1 + w}k^2\delta - k^2\Psi = \frac{a^2\partial^i Q_i}{(1 + w)\bar{\rho}} - \frac{aQ_0}{\bar{\rho}}\theta \tag{6.22}$$

From this point on, the bar over background quantities will be dropped.

6.5 Perturbed Einstein equations

The Einstein equations will also be modified, and we can relate the first order changes in the metric (culminating in the Einstein tensor $G^{\mu\nu}$) to the perturbations in the energy-momentum tensor.

To first order, the perturbed Einstein equations are

$$k^2\phi + 3\frac{\dot{a}}{a}\left(\dot{\phi} + \frac{\dot{a}}{a}\psi\right) = -4\pi G a^2 \delta\rho \tag{6.23}$$

$$k^2\left(\dot{\phi} + \frac{\dot{a}}{a}\psi\right) = -4\pi G a^2 \left(\rho + P\right)\theta\,, \tag{6.24}$$

$$\ddot{\phi} + \frac{\dot{a}}{a}(\dot{\psi} + 2\dot{\phi}) + \left(2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}\right)\psi + \frac{k^2}{3}(\phi - \psi) = -\frac{4\pi}{3}\delta P\,, \tag{6.25}$$

$$k^2(\phi - \psi) = 0\,. \tag{6.26}$$

where the final equation has assumed a lack of any anisotropic stress (see [103] for full details on cosmological perturbation theory).

6.6 Energy-momentum balance

Energy-momentum conservation for fluid A implies

$$\nabla_\mu T^\mu_{\ \nu} = Q_\nu,$$

and conservation for the entire system requires

$$\sum_A Q^\mu_A = 0. \tag{6.27}$$

The peculiar velocity potential v_A is related to θ_A via $\theta_A = -k^2 v_A$. The total velocity potential is given by

$$\left(\rho + P\right) v = \sum_A \left(\rho_A + P_A\right) v_A \tag{6.28}$$

A general energy-momentum exchange vector can be split as

$$Q^A_0 = Q_A\left(1 + \psi\right) + \delta Q_A \tag{6.29}$$

$$Q^A_i = -a\partial_i\left(f_A + Q_A v\right). \tag{6.30}$$

Here, f_A is an intrinsic momentum potential, while the term proportional to the total velocity potential is due to velocity transport along the total velocity. Conservation of energy requires

$$0 = \sum Q_A = \sum \delta Q_A = \sum f_A \tag{6.31}$$

6.7 Sound speed

The sound speed c_{sA} of fluid or scalar field A is the speed at which pressure fluctuations propagate in the rest frame of A,

$$c_{sA}^2 = \frac{\delta P_A}{\delta \rho_A}\Big|_{rf} \tag{6.32}$$

Meanwhile, the "adiabatic sound speed" for the medium A is defined as [144]

$$c_{aA}^2 = \frac{P_A'}{\rho_A'} = w_A + \frac{w_A'}{\rho_A'/\rho_A} \tag{6.33}$$

The adiabatic sound speed is the sound speed a fluid would have, assuming that there are no non-adiabatic perturbations present. For barotropic fluids, the sound speed and the adiabatic sound speed are identical. For fluids with a constant, negative w, this obviously leads to an imaginary sound speed, which produces instabilities in the dark energy. Most authors instead adopt the scalar field value;

$$c_{sx}^2 = 1, \qquad c_{ax}^2 = w_x < 0. \tag{6.34}$$

This introduces a non-adiabatic pressure perturbation (see [144] for details on the short derivation),

$$\delta P_A = c_{sA}^2 \delta\rho_A + \left(c_{sA}^2 - c_{aA}^2\right)\left[3\mathcal{H}(1+w_A)\rho_A - aQ_A\right]\frac{\theta_A}{k^2} \tag{6.35}$$

By adopting a scalar-field sound-speed, the sound speed of the fluid is no longer that of the adiabatic sound-speed. Non-adiabatic pressure perturbations have been introduced. We will return to this point in Section 7.2.3 when we come to solve the equations for a specific scenario.

Substituting equations (6.29) and (6.30) for the energy exchange four-vector, along with equation (6.35) for the pressure perturbation, into equations (6.18) and (6.22), then the perturbation equations for fluid A take the most general form:

$$\begin{aligned}\delta_A' + 3\mathcal{H}(c_{sA}^2 - w_A)\delta_A + (1+w_A)\theta_A + 3\mathcal{H}\left[3\mathcal{H}(1+w_A)(c_{sA}^2 - w_A) + w_A'\right]\frac{\theta_A}{k^2} \\ -3(1+w_A)\Phi' = \frac{aQ_A}{\rho_A}\left[\Phi - \delta_A + 3\mathcal{H}\left(c_{sA}^2 - w_A\right)\frac{\theta_A}{k^2}\right] + \frac{a}{\rho_A}\delta Q_A\end{aligned} \tag{6.36}$$

$$\begin{aligned}\theta_A' + \mathcal{H}(1-3c_{sA}^2)\theta_A - \frac{c_{sA}^2}{(1+w_A)}k^2\delta_A - k^2\Psi = \frac{aQ_A}{(1+w_A)\rho_A}\left[\theta - (1+c_{sA}^2)\theta_A\right] \\ -\frac{a}{(1+w_A)\rho_A}k^2 f_A.\end{aligned} \tag{6.37}$$

6.8 Final notes

Discovering how structure evolves in the universe at a linear level is tantamount to solving equations (6.36)–(6.37), supplemented by the perturbed Einstein equations, (6.23)–(6.26), and of course the Friedmann equation to provide $\mathcal{H}$. For multiple fluids, these coupled equations can become difficult to solve.

As a simplification, and to shed some insight on the physics behind the mathematics, we will consider the sub-horizon ($k/\mathcal{H} \gg 1$) limit, and a pure matter (de Sitter) universe.

The equation of motion for δ simplifies to,

$$\delta' + 3\mathcal{H}\delta + \theta = 3\phi' \,. \tag{6.38}$$

Meanwhile, one of the perturbed Einstein equations simplifies to what amounts to Poisson's equation in comoving coordinates,

$$-k^2\psi = 4\pi G a^2 \rho\, \delta \,. \tag{6.39}$$

Combining with the equation for θ, and neglecting time derivatives of $\phi = \psi$ (from the last perturbed Einstein equation) leads to a second order equation,

$$\delta'' + \mathcal{H}\delta = -k^2\psi \,. \tag{6.40}$$

The growth of structure is driven by gravity (the right hand side), but suppressed by the expansion of the universe.

In the matter dominated era, the solution is $\delta_c \propto \tau^2 \propto a$; density perturbations grow linearly with the expansion of the universe.

The following two Chapters will carry out much the same analysis, but in a universe where energy exchange between dark matter and dark energy is allowed.

Chapter 7

On the Large-Scale Instability in Interacting Dark Energy and Dark Matter Fluids

By far the simplest model of dark energy is Einstein's cosmological constant, Λ. The cosmological constant(Λ) and cold dark matter (CDM) model, with values of today's density parameter for the dark energy $\Omega_x \approx 0.7$ and dark matter $\Omega_m \approx 0.3$ is the current prevailing paradigm. But while consistent with observational constraints, the standard model is in many ways unsatisfactory. One such example is the 'coincidence problem': why are the energy densities in the dark energy and dark matter comparable today, when the redshift dependence of each is so different?

Motivated to explain the coincidence problem while deviating as little as possible from the successful ΛCDM model, a coupling between dark energy and dark matter has often been considered. An energy exchange modifies the background evolution of the dark sector, and explaining the coincidence problem can be reduced to tuning a coupling parameter to an appropriate value.

The coupling enters via the continuity equations. With energy exchange rate Q between the dark energy (subscript x) and the cold dark matter (subscript c), the dark

energy obeys the continuity equation in conformal time

$$\rho_x' + 3\mathcal{H}(1 + w_x)\rho_x = -Q, \tag{7.1}$$

while the dark matter obeys

$$\rho_c' + 3\mathcal{H}\rho_c = Q. \tag{7.2}$$

Here we have introduced the equation of state parameter w_A that gives the ratio of the pressure P_A to the energy density ρ_A of a fluid,

$$w_A = \frac{P_A}{\rho_A}. \tag{7.3}$$

We have also used $\mathcal{H} = aH$, where $a(t)$ is the expansion scale-factor and H the Hubble parameter, and primes to indicate derivatives with respect to conformal time. Acceleration of the expansion rate requires the energy density of the universe to dominated by a fluid with an effective equation of state parameter $w_{\text{eff}} < -1/3$. We do not allow the phantom case of $w < -1$ in this chapter. Simple solutions for the background exist for couplings of the form $Q = \alpha\mathcal{H}\rho_x + \beta\mathcal{H}\rho_c$. These were initially investigated by Chimento [38] and then expanded upon by Barrow and Clifton [17], who provided general solutions for any cosmology with two components exchanging energy in such a fashion, provided the components were modelled as cosmological fluids with constant w. Quartin et al. [121] examined the observational constraints upon such a class of models, significantly limiting the available parameter space. Again, the equation of state parameter was treated as fixed. Non-zero values of β were found to reduce the required fine-tuning of the initial energy density, as well as increase the observationally allowed values of α [121].

A coupling would influence more than just the background dynamics of the universe. In particular, the growth of perturbations in the coupled fluids would be affected. Recent work by Valiviita et al. [144] has shown that couplings of the simple form described above, with constant w, exhibit extremely rapid growth of dark energy fluctuations on super-horizon scales in the early universe. In fact, the perturbations in the dark energy become unstable for any model with non-zero β, no matter how small this parameter is made. While this would appear to rule out all couplings of the above form and with constant w, the explicit examples in [144] included no cases where the interaction rate was proportional to the density of dark energy and not of the dark matter, i.e. with $\beta = 0$ and $\alpha \neq 0$. Here we look at just such a scenario.

7.1 Background evolution

Friedmann's equation relates the evolution of the scale-factor $a(t)$ to the background energy density ρ. We make use of conformal time, τ, which is related to cosmic time via $dt = ad\tau$. Overdots indicate derivatives with respect to conformal time.

Friedmann's equation reads

$$\mathcal{H}^2 \equiv \left(\frac{a'}{a}\right)^2 = \frac{8\pi G}{3}\rho a^2. \tag{7.4}$$

With the choice of $Q = \alpha\mathcal{H}\rho_x$, the continuity equations can be solved to yield [17, 38]

$$\rho_x = \rho_{x,0}\, a^{-(3(1+w)+\alpha)}, \tag{7.5}$$

$$\rho_c = -\frac{\alpha\rho_{x,0}}{3w+\alpha}\, a^{-(3(1+w)+\alpha)} + \left(\rho_{c,0} + \frac{\alpha\rho_{x,0}}{3w+\alpha}\right) a^{-3}.$$

We follow the standard notation where a subscript zero indicates today's value. We normalise the scale-factor so that $a_0 = 1$. The ratio of dark energy to dark matter density r can then be written

$$\frac{1}{r} = \frac{\rho_c}{\rho_x} = \left(\frac{\rho_{c,0}}{\rho_{x,0}} + \frac{\alpha}{3w+\alpha}\right) a^{3w+\alpha} - \frac{\alpha}{3w+\alpha} \tag{7.6}$$

With $|3w| < \alpha$, the dark energy and dark matter approach a constant ratio as the universe expands. The coincidence problem can be said to be solved if this ratio is of order unity, but this requires a value of α already observationally excluded [121]. Nevertheless, as argued in [121], non-zero values of α can still be said to alleviate the problem. We restrict ourselves to positive values of α.

7.2 Perturbed FRW cosmology

We assume a flat FRW cosmology and work in Newtonian gauge,

$$-ds^2 = dt^2(1+2\Psi) - a^2(1-2\Phi)\delta_{ij}dx^i dx^j, \tag{7.7}$$

with metric signature $(+,-,-,-)$. We work in Fourier space, using comoving Fourier wave-vectors $k^i = k_i$, so that $\partial_i\partial^i \to -k^2/a^2$.

The four-velocity of fluid A is given by

$$U^\mu_{(A)} = \left((1-\Psi), a^{-1}v^i_{(A)}\right). \tag{7.8}$$

The peculiar velocity three-vector $v^i = v_i$ are small. We define the velocity perturbation $\theta \equiv \partial_i v^i$.

7.2.1 Energy-momentum tensors

The energy-momentum tensor for fluid A is given by:

$$T^{\mu(A)}_{\ \ \nu} = \left(\rho^{(A)} + P^{(A)}\right) U^{(A)\mu} U^{(A)}_{\nu} - \delta^{\mu}_{\nu} P^{(A)}. \tag{7.9}$$

The total energy-momentum tensor is simply the sum of the components,

$$T^{\mu}_{\ \ \nu} = \sum_{A} T^{\mu(A)}_{\ \ \nu}. \tag{7.10}$$

We define the density perturbation in fluid δ_A using $\rho_A \equiv (1+\delta_A)\bar{\rho}_A$. An overbar denotes the background quantity, though we will usually leave this implicit.

Energy and momentum conservation for fluid A implies

$$\nabla_{\mu} T^{\mu(A)}_{\ \ \nu} = Q^{(A)}_{\nu},$$

and conservation for the entire system requires

$$\sum_{A} Q^{\mu}_{(A)} = 0. \tag{7.11}$$

The four-vector $Q^{\mu}_{(A)}$ governs the energy exchange between components, and it is to this we now turn our attention.

7.2.2 Covariant energy exchange

The energy exchange in the background does not determine a fully covariant form of energy exchange [88, 144]. Instead, an energy exchange four-vector must be specified. We adopt the approach of [144] and consider two scenarios; aligning the four-vector with the dark energy four-velocity,

$$Q^{\mu}_{(A)} = Q_A U^{\mu}_x, \tag{7.12}$$

or with the four-vector of the dark matter four-velocity,

$$Q^{\mu}_{(A)} = Q_A U^{\mu}_c. \tag{7.13}$$

These choices produce slightly different outcomes, and the differences are noted as we proceed.

To produce the desired changes to the continuity equations, we see that $aQ_c = -aQ_x = \alpha\mathcal{H}\rho_x$ in both cases. We also make the common assumption that $\alpha\mathcal{H}$ is an approximation to an interaction rate that has no spatial dependence. We therefore

perturb only ρ_x, not $\mathcal{H}$ in the coupling, though this means the energy exchange is no longer covariant (to be so, the time derivative of a should include the effect of the perturbations to the time part of the metric; $\mathcal{H}$ is no longer precisely scalar once perturbations are included). Thus the interaction rate depends purely on the large-scale cosmic time, and not on the proper time of local observers.

7.2.3 Sound speed of dark energy

The speed of sound of a fluid or scalar field A is denoted by c_{sA}. For a barotropic fluid with a constant value of w_A, then $c_{sA}^2 = w_A$. This leads to an imaginary speed of sound for the dark energy ($c_{sx}^2 = w_x < 0$). An imaginary sound speed leads to instabilities in the dark energy; the problem is commonly remedied by imposing a real sound speed by hand. A common choice (and the one we make here) is the scalar field value of $c_{sx} = 1$.

This choice leads to an intrinsic non-adiabatic pressure perturbation in the dark energy. This contains a term, highlighted recently in [144], that arises due to the coupling between dark energy and dark matter. We include this term, and refer the interested reader to [144].

7.2.4 Perturbation equations of motion

Conservation of the energy-momentum tensor, combined with results of the previous sections and our choice of energy exchange four-vector, implies the following. For the dark energy density perturbation:

$$\begin{aligned} \delta_x' &+ 3\mathcal{H}(1-w_x)\delta_x + (1+w_x)\theta_x + 9\mathcal{H}^2(1-w_x^2)\frac{\theta_x}{k^2} \\ &- 3(1+w_x)\Phi' = -\alpha\mathcal{H}\left[\Psi + 3\mathcal{H}(1-w_x)\frac{\theta_x}{k^2}\right]. \end{aligned} \tag{7.14}$$

For the dark energy velocity perturbation, the right-hand side differs slightly depending on our choice of energy exchange four-vector.

$$\theta_x' - 2\mathcal{H}\theta_x - \frac{k^2}{1+w_x}\delta_x - k^2\Psi = \frac{(1+b)\alpha\mathcal{H}}{1+w_x}\theta_x,$$

where

$$b = \begin{cases} 0 & \text{if } Q^\mu_{(A)} = Q_A U^\mu_x, \\ 1 & \text{if } Q^\mu_{(A)} = Q_A U^\mu_c. \end{cases} \tag{7.15}$$

For the dark matter, the density perturbation obeys

$$\delta_c' + \theta_c - 3\Phi' = \alpha\mathcal{H}\frac{\rho_x}{\rho_c}\left[\delta_x - \delta_c\right], \tag{7.16}$$

while the velocity perturbation is governed by

$$\theta_c' + \theta_c\mathcal{H} - k^2\Psi = (1-b)\alpha\mathcal{H}\frac{\rho_x}{\rho_c}\left[\theta_x - \theta_c\right]. \tag{7.17}$$

The perturbed Einstein equations are well known, and we do not reproduce them here. They can be found in [103], whose notation for the scalar metric perturbations we share.

7.3 Initial conditions in the early radiation era

In [144] it was shown that models with $\beta \neq 0$ suffered from an early time large-scale instability no matter how small the value of β. This was driven by a term proportional to β on the right-hand side of equation (7.15). A term proportional to α also exists, which can be large if w is close to -1 or α is made very large. In this section we examine how large this term needs to be to cause the non-adiabatic mode to be a growing one.

We consider super-horizon scales ($k/\mathcal{H} \ll 1$) and assume initial conditions that would be adiabatic in the uncoupled case (though we have not made any attempt to determine if either mode corresponds to an entropy perturbation). The gravitational potentials are dominated by fluctuations in the dominant fluid (radiation or matter). The well known result is that $\Phi \propto \Psi =$ constant. The constant of proportionality in the radiation era is determined by the anisotropic stress generated by the neutrinos. In the absence of neutrinos or in the matter dominated era, the potentials are equal. These assumptions will be invalid only if perturbations in the dark energy are large enough to influence the gravitational potentials. As the dark energy has a very low background density in the radiation era, this can only happen if δ_x grows extremely large.

Neglecting time derivatives of the gravitational potential, and keeping only leading order terms in $k/\mathcal{H}$, the dark energy equations (7.14) – (7.15) can be combined into a second order equation:

$$\begin{aligned}
\delta_x'' &+ \mathcal{H}\left(1 - 3w - \frac{(1+b)\alpha}{1+w} - 2\frac{\mathcal{H}'}{\mathcal{H}^2}\right)\delta_x' \\
&+ 3\left(\mathcal{H}^2\left(1 - b\frac{\alpha}{1+w}\right) - \mathcal{H}'\right)(1-w)\,\delta_x \\
&= \left(A\mathcal{H}^2 + B\mathcal{H}'\right)\Psi,
\end{aligned} \tag{7.18}$$

The constants A and B have values unimportant for our analysis.

In the radiation era, $\mathcal{H} = \tau^{-1}$. An obvious solution is therefore: $\delta_x \propto \Psi =$ constant. This would be the adiabatic mode in the uncoupled case. To find the remaining solutions, we take this solution as dominant and look for how the others grow given this case. We assume Ψ is approximately constant and define a new variable $\hat{\delta}_x = \delta_x + C\Psi$, with the constant C chosen such that the right-hand side of (7.18) is equal to zero. In the radiation dominated era, we can then write:

$$\tau^2\hat{\delta}_x'' \quad + \quad \left(3 - 3w - \frac{(1+b)\alpha}{1+w}\right)\tau\hat{\delta}_x' + 3(1-w)\left(2 - b\frac{\alpha}{1+w}\right)\hat{\delta}_x = 0 \qquad (7.19)$$

When $b = 1$, equation (7.19) becomes formally the same equation found by He et al.[72], despite the differing assumptions made about the physics involved. In their investigation of perturbations given a background coupling of the form $Q = \alpha\mathcal{H}\rho_x$, they choose to set the net momentum exchange to zero ($Q^i_{(A)} = 0$), in contrast to our adoption of the form of momentum exchange used in [144]. The differences between the $b = 1$ choice of momentum exchange and zero net momentum exchange arise in the equations for the dark matter perturbations, which are not used in the above analysis, nor in the analysis by He et al. [72] This leads to the same behaviour of dark energy perturbations. This is not true when $b = 0$, and can result in different behaviour (oscillatory or non-oscillatory) for the same choice of parameters (see the remainder of this section). Note also that the simplifying assumptions, and the justifications, made in [72] differ to those made here: we have neglected terms that will be small due to choice of intial conditions, and simplified the result by assuming the mode that would be adiabatic in the uncoupled case is dominant. In [72], terms are instead neglected that are found to be small from a numerical analysis.

Solutions of equation (7.19) are power laws, $\hat{\delta}_x \propto \tau^{n_\pm}$. The index is given by:

$$n_\pm = \frac{\Gamma}{2(1+w)} \pm \frac{\sqrt{\Delta}}{2(1+w)}, \qquad (7.20)$$

where we follow the notation of [72] and have defined the quantities

$$\Gamma = 3w^2 + w + (1+b)\alpha - 2, \qquad (7.21)$$

and

$$\begin{aligned}\Delta &= 9w^4 + 30w^3 + (13 - 6(b-1)\alpha)w^2 \\ &+ 2w\left[(1+b)\alpha - 14\right] + 4(2b-1)\alpha + (1+b)^2\alpha^2 - 20. \end{aligned} \qquad (7.22)$$

In the limit of w very close to -1 (and assuming α is reasonably small), we can expand as a series in $(1+w)$,

$$\frac{\Gamma}{2(1+w)} \approx -5/2 + \frac{(1+b)\alpha}{2(1+w)} \quad + \quad \frac{3(1+w)}{2} + \mathcal{O}(1+w)^2, \qquad (7.23)$$

$$\begin{aligned}\Delta &\approx (1+3b)\alpha^2 + 2(7b-5)(1+w)\alpha \\ &+ (6(1-b)\alpha - 23)(1+w)^2 + \mathcal{O}(1+w)^3. \end{aligned} \tag{7.24}$$

When $\alpha = 0$, the non-adiabatic mode is decaying. But when the coupling is switched on, the second term in Γ can become very large, resulting in $n_+ \gg 1$. For a range of α and w, which is much larger in the $b = 0$ case, oscillatory behaviour can also result (due to Δ becoming negative). The instability means these coupled models suffer from all the problems outlined in [144] for $\beta \neq 0$ models, unless the value of α is made small enough. The closer w is to -1, the smaller α must be made to avoid the instability. This is in contrast to $\beta \neq 0$ models, which are unstable no matter how small the parameter β is made.

In the matter dominated era, we can carry out the same procedure, this time with $\mathcal{H} = 2\tau^{-1}$. We find,

$$\frac{\Gamma}{2(1+w)} \approx -9/2 + \frac{(1+b)\alpha}{1+w} + 3(1+w) + \mathcal{O}(1+w)^2, \tag{7.25}$$

and

$$\begin{aligned}\Delta &\approx 4(1+3b)\alpha^2 + 12(5b-3)\alpha(1+w) \\ &+ (24(1-b)\alpha - 62)(1+w)^2 + \mathcal{O}(1+w)^3. \end{aligned} \tag{7.26}$$

Once again, the second term in Γ can result in a rapidly growing dark energy fluctuation.

We have solved equations (7.14) – (7.17) numerically in the matter dominated regime (Figure 7.1), where we need not worry about the radiation fluid and its perturbations. The analytical agreement is excellent until the mode leaves the horizon ($k\tau \sim 1$). Numerically we see that when this happens the mode begins to oscillate with a growing amplitude.

7.4 Sub-horizon evolution in the matter and radiation dominated eras

In the sub-horizon limit, $\mathcal{H}^2/k^2 \ll 1$, equation (7.14) yields,

$$\delta_x' + 3\mathcal{H}(1-w_x)\delta_x + (1+w_x)\theta_x - 3(1+w_x)\Phi' = 0. \tag{7.27}$$

Note the two terms on the right-hand side of equation (7.14) scale as $\mathcal{H}^2/k^2$. As these are the only two terms containing the coupling parameter α, the simplified equation above does not contain the coupling parameter.

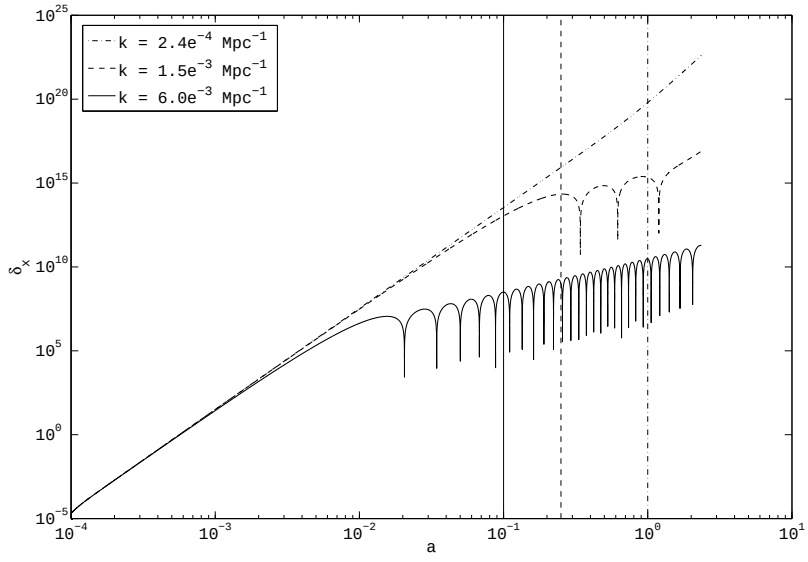

Figure 7.1: The evolution of the dark energy outside the horizon in a matter dominated universe, for modes of three different scales. We take $\alpha = 0.08$ and $w = -0.98$, and $b = 1$. The agreement with the analytical approximation is excellent until the mode begins to leave the horizon ($k\tau \sim 1$). Vertical lines indicate when this occurs for each mode.

One of the perturbed Einstein equations simplifies to Poisson's equation in comoving coordinates,

$$-k^2\Psi = 4\pi G a^2 \left(\rho_x \delta_x + \rho_c \delta_c\right). \tag{7.28}$$

Without the coupling, the dark energy perturbations are significantly suppressed on small-scales in comparison to dark matter perturbations, primarily due to its large speed of sound [21]. The coupling does nothing to alter this fact unless the right-hand side of equation (7.15) makes a significant contribution. If the early time instability has been avoided this cannot be the case, as $\alpha/(1+w)$ will be small. Thus it is reasonable to expect the dark energy to remain suppressed on sub-horizon scales. We therefore neglect dark energy perturbations for the remainder of this section.

By combining equations (7.16) and (7.17), we eliminate θ_c and find a second-order equation for the growth of the matter density perturbation. From the above argument, we have neglected dark energy perturbations. Keeping only the dominant gravitational terms,

$$\begin{aligned} \delta_c'' &+ \mathcal{H}\left(1 + 2\alpha\frac{\rho_x}{\rho_c}\right)\delta_c' \\ &+ \alpha\frac{\rho_x}{\rho_c}\left(\mathcal{H}' - \mathcal{H}^2(\alpha + 3w - 1)\right)\delta_c = -k^2\Psi. \end{aligned} \tag{7.29}$$

We note that in the limit of $\alpha \to 0$, this reduces to the standard growth equation, with

the well known growing mode $\delta_c \propto \tau^2$ in both matter and radiation eras. The additional terms are proportional to αr (recall r is the ratio of dark energy to dark matter). In the matter dominated regime, then $\alpha r \ll 1$, and these terms will be negligible. Even when $r \sim 1$, the terms will be suppressed by the size of α, which will be small itself. The dominant effect causing a deviation from standard linear growth of structure in the matter dominated regime will therefore be, as in an uncoupled cosmology, the influence of the dark energy upon the expansion rate. The growth of structure in a coupled model can therefore be treated in the matter dominated regime simply as an uncoupled model with an effective dark energy equation of state parameter $w_{\rm eff} = w + \alpha/3$. This will cease to be true only when the background energy density of matter is no longer well approximated by its usual $\rho_c \propto a^{-3}$ dependence, and the late time scaling behaviour becomes apparent.

7.5 Sub-horizon evolution in the dark energy dominated era

The coupling between dark energy and dark matter eventually leads to a constant ratio between the two dark components. With a small value of α, the dark energy still dominates. We consider the evolution of structure once this equilibrium has been reached.

Friedmann's equation solves to yield

$$\mathcal{H} = 2(\alpha + 3w + 1)^{-1}\eta^{-1}, \tag{7.30}$$

with the new time variable $\eta = \tau - \tau_\infty$. Note that as η increases (τ decreases and approaches τ_∞), the scale-factor increases. The constant of integration, τ_∞, is the radius of the de Sitter event-horizon in the uncoupled case with a cosmological constant. The growth equation can then be written as,

$$\begin{aligned} \eta^2\delta_c'' &+ \eta\frac{2-12w-4\alpha}{\alpha+3w+1}\delta_c' \\ &+ 2\frac{(3\alpha+9w-1)(3w+\alpha)+\alpha w}{(\alpha+3w+1)^2}\delta_c = 0. \end{aligned} \tag{7.31}$$

This admits power law solutions, $\delta_c \propto \eta^m$, where

$$m = \frac{5}{2} - \frac{3}{1+3w+\alpha} \pm \frac{1}{2}\sqrt{1 - \frac{8\alpha}{w(1+3w+\alpha)^2}}. \tag{7.32}$$

In the range of α and w relevant to the problem, then $m > 0$. Recalling that η decreases with increasing scale-factor, we see that the universe becomes steadily more

homogeneous as it expands. We interpret this to be a combination of two effects. The first is the accelerating expansion, which slows and (without the coupling) eventually stops structure formation. This occurs, for example, in ΛCDM cosmology when the cosmological constant becomes dominant. The second effect is that dark energy is constantly being transformed into dark matter, via the coupling. As the rate is proportional to the density of the dark energy, and the dark energy density is essentially uniform, new dark matter is also created uniformally. This rising 'background' of dark matter reduces the relative value of the fluctuations, reducing δ_c.

We have also investigated both numerically and analytically the extreme late time behaviour, where $k\eta \ll 1$ and the modes can be thought of as leaving the horizon. We find the tend toward homogeneity continues, but with a much milder rate of decay.

7.6 Conclusions

We have shown that constant w models with the same form of energy-momentum exchange considered by [144] suffer from an instability with $\alpha \neq 0$, even if $\beta = 0$. However the instabilities in these models are not as severe as those facing models with $\beta \neq 0$. There is at least some non-trivial region of parameter space where the instability can be avoided, although the value of α is now constrained both from background observables [121] and from stability requirements to be extremely small. Despite this, any non-zero value of α will lead to a late-time scaling regime, alleviating (even if not solving) the coincidence problem. It is unfortunate that with α constrained to such small values, we find any observable trace upon the growth of CDM structure will be negligible. Detecting a coupling of this form from measurements of large-scale structure is extremely doubtful, even with the precision promised by future experiments.

We have said nothing up to this point of models of dark energy with a variable equation of state parameter, such as scalar-field (quintessence) models. The same caveats in [144] apply here. Much of the above analysis will not apply in variable w models, although some parameterisations such as the often used $w = w_0 + (1-a)w_a$ lead to fixed w over large periods of time. Our analysis will apply during those epochs of constant w. We refer interested readers to recent work on quintessence with couplings of this or similar form (such as recent work [39, 47]).

The future decay of dark matter fluctuations is an interesting result. It implies observers today find themselves close to the time of maximum inhomogeneity. The more the coincidence problem is alleviated, the closer to the late-time scaling regime

today becomes, and thus the closer to the peak of inhomogeneity. Without the coupling, observers find themselves at the time of the end of structure growth. The root cause in both cases is the acceleration of the universe only beginning today. Our position as apparently privileged observers in this fashion remains difficult to explain in any satisfactory way.

Chapter 8

Unmodified Gravity

8.1 Introduction

One of the major goals of future cosmological studies is to categorise dark energy as one of three candidates: a cosmological constant, a physical fluid, or simply some manifestation of new gravitational physics. Geometric measurements such as supernovae and baryon acoustic oscillations will scrutinise the null hypothesis of a cosmological constant. However, a physical fluid and modified gravity could both reproduce almost arbitrary expansion histories, so distinguishing them requires a study of the growth of structure. It is the validity of this diagnostic step that provides the focus of the present work.

Anisotropic stress in the dark energy fluid may mimic metric theories of gravity, as demonstrated by Kunz & Sapone [92] and more generally by Hu [75]. Here we take this one step further: even *without* anisotropic stress, it can be shown that significant deviations in structure growth are achievable. The only requirement is that some interaction should exist between dark matter and the smooth dark energy component, permitting the transfer of energy.

Any change in the dark energy density is conventionally attributed to the equation of state $w(z)$, which dictates the adiabatic behaviour of a physical fluid. But density evolution may also arise from non-gravitational interactions with other fluids, thereby violating adiabaticity. Cosmologies with energy exchange have been extensively studied

in the literature [5, 17, 144, 143, 39, 78, 157, 12, 94], and generate an expansion history that is fully reproducible by a single inert scalar field whose evolution matches the effective equation of state $w_{\rm eff}(z)$. The degeneracy between these two models may be broken by studying the growth of structure, which is disrupted by the evolving matter density [72]. Yet it is this same test that would conventionally be used to identify modifications to gravity.

In Section 8.2 we begin by reviewing a simple case of dark energy decaying into a form of dark matter. The evolution of perturbations are quantified in Section 8.3, before exploring the observational consequences in Section 8.4.

8.2 Energy exchange

When speculating on possible interactions amongst the lesser known constituents of our Universe, we are spanning a remarkably broad class of models, with potentially numerous degrees of freedom. This could extend to new regimes of dark physics, such as dark matter particles spontaneously decaying into a relativistic dark species such as neutrinos or massless particles, e.g. the dark photons speculated by Ackerman et al [2]. This particular example would be compatible with current observations provided the dark matter particle is sufficiently massive, and the dark fine-structure constant is sufficiently small. For the remainder of this work, we shall focus on the case of dark energy decaying into a form of dark matter, and explore the observational consequences.

As an illustrative example, we study the simple case of a cosmological (almost) constant that decays into a homogeneous form of dark matter. The selection $w = -1$ for the dark energy fluid bypasses the various instability issues highlighted in previous work [78, 144, 39]. Instead we choose to focus on the behaviour of the dark matter perturbations. The evolution of the mean matter density ρ_m is dictated by the conservation equations

$$\rho'_\Lambda = -Q, \tag{8.1}$$

$$\rho'_m + 3\mathcal{H}\rho_m = Q, \tag{8.2}$$

where the prime denotes a derivative with respect to conformal time, and $\mathcal{H} \equiv a'/a = \dot{a}$ is the conformal Hubble parameter. The interaction parameter Q controls the rate of energy transfer, and is generally considered to be a function of either the dark matter or dark energy density.

The central result of this chapter will be to demonstrate that models of interacting dark energy modify the growth of large scale structure, with an explicit example that exhibits a constant decrement c, such that

$$\begin{aligned} f &\equiv \frac{d\ln\delta}{d\ln a} \\ &= \Omega_m^\gamma - c, \end{aligned} \tag{8.3}$$

thus providing another mechanism for anomalous growth, aside from modified gravity or anisotropic stress. The magnitude of c is determined by both the nature and strength of the interaction.

We adopt an empirical modification to the density evolution, which reproduces the functional form for a number of models such as interacting quintessence [157, 147], given by

$$\rho_m = \rho_{m0} a^{-3+\epsilon}, \tag{8.4}$$

where the parameter $\epsilon \ll 1$ dictates the rate of energy transfer. This particular parameterisation is advantageous in providing the necessary scaling of the dark energy density to help resolve the naturalness and coincidence problems. The dark energy density maintains a magnitude comparable to the dark matter density, although it does not help explain why the transition to dark energy domination has occurred at recent times. Furthermore, the simple form of (8.4) allows significant progress to be made analytically. In this case the evolution of the dark energy density is, for a flat universe, given by

$$\rho_\Lambda = \rho_{\Lambda 0} + \frac{\epsilon}{3-\epsilon}\rho_{m0}\left[a^{-3+\epsilon} - 1\right]. \tag{8.5}$$

At high redshifts, this leads to the tracking behaviour characterised by

$$\Omega_\Lambda(a \to 0) = \frac{\epsilon}{3}, \tag{8.6}$$

until we approach the era of radiation domination. Thus in effect this model adjusts the dark energy decay lifetime with epoch; for a constant lifetime, we would expect $\rho_\Lambda \to$ constant at high z and $\rho_\Lambda \to 0$ at late times. Our analysis treats the parameter ϵ as a constant, though it remains valid provided ϵ varies sufficiently slowly, satisfying the condition $d\epsilon/\,d\ln a \ll \epsilon$.

Given the dark matter mass evolution (8.4), one can see a direct relation between our phenomenological approach, and that of a coupled scalar field as outlined in [5].

$$\epsilon_{\rm eff} = \frac{-\int Q(\phi)\, d\phi}{\ln a}. \tag{8.7}$$

8.3 Growth of structure

Any form of non-gravitational interaction in the dark sector may be expected to impact upon the growth of structure. As we shall see, a decaying form of dark energy generates three distinct mechanisms for slowing the rate of structure growth. We quantify each of these in turn, starting with the modification to the expansion history $H(z)$.

8.3.1 Background dynamics

Here we generalise the derivation of linear structure growth by Linder & Cahn [101] to incorporate the different background evolution of both dark energy and dark matter, $\rho_m \propto a^{-3+\epsilon}$. Including energy exchange at the background level *only* is insufficient, as we will show, but it is instructive to identify the contribution this yields in the following derivation of $f(a) \equiv d\ln\delta/d\ln a$. Starting from the differential equation

$$\frac{d^2\delta}{dt^2} + 2H(a)\frac{d\delta}{dt} - 4\pi\rho_m\delta = 0, \tag{8.8}$$

which may be rewritten as

$$\frac{df}{d\ln a} + \frac{1}{2}\frac{d\ln H^2}{d\ln a}\left(f - \frac{1}{2}\right) + f(f+2) - \frac{3}{2}\Omega_m(a) = 0\,, \tag{8.9}$$

and utilising

$$H^2/H_0^2 \simeq \Omega_m a^{-3+\epsilon}\left[1 + \Omega_\Lambda(a)/\Omega_m(a)\right]\,, \tag{8.10}$$

yields, to first order in $(f(a) - 1)$ [101],

$$f(a) \approx 1 - \frac{\epsilon}{5} - \frac{1}{2}\Omega_\Lambda(a)\left(1 - \frac{\epsilon}{5}\right) + \left[1 - \frac{1}{2}\Omega_\Lambda(a)\right] I(a)\,, \tag{8.11}$$

where

$$I(a) = -\frac{1}{4}a^{-(5+\epsilon)/2}\int_0^a \frac{da'}{a'}(a')^{(5+\epsilon)/2}\Omega_\Lambda(a')\,. \tag{8.12}$$

Before evaluating this integral, we first need to consider the scaling behaviour of dark matter and dark energy at early times. To proceed, we simply recast the dark energy density (8.5) in the form

$$\Omega_\Lambda(a) = \Omega_1 + \Omega_2(a) + O\left(\epsilon^2\right) + O\left(\Omega_\Lambda^2(a)\right)\,, \tag{8.13}$$

where Ω_1 is the constant component, and Ω_2 is the collection of terms that follow a power law:

$$\begin{aligned}\Omega_1 &= \frac{\epsilon}{3};\\ \Omega_2 &= \left(\Omega_{\Lambda 0} - \frac{\epsilon}{3}\right) a^{3-\epsilon}\,,\end{aligned} \tag{8.14}$$

provided we assume the contribution from radiation is negligible.

Using these redshift dependencies in $I(a)$, then to first order in deviations from matter domination we find:

$$I(a) = -\frac{1}{4}\left[\frac{2}{5+\epsilon}\Omega_1 + \frac{2}{11-3\epsilon}\Omega_2(a)\right]\,. \tag{8.15}$$

Writing $\Omega_2(a) = \Omega_\Lambda(a) - \Omega_1$ leaves us with

$$I(a) = -\frac{\epsilon}{30} + \frac{\epsilon}{66} - \frac{\Omega_\Lambda(a)}{22} + O\left(\epsilon^2\right) + O\left(\Omega_\Lambda^2(a)\right)\,. \tag{8.16}$$

Substituting our solution back into (8.11) gives:

$$f(a) \simeq 1 - \left(\frac{6}{11} - \frac{6}{55}\epsilon\right)\Omega_\Lambda(a) - \frac{12}{55}\epsilon \tag{8.17}$$

Of these two new terms involving ϵ, the second is of greater importance. This contrasts with the conventional approximation for $f(a)$ [146, 100] given by

$$\begin{aligned}f(a) &= \Omega_m^\gamma(a)\\ &\simeq 1 - \gamma\Omega_\Lambda(a)\,.\end{aligned} \tag{8.18}$$

If we insist on maintaining the definition $\gamma \approx [1 - f(a)]\,/\Omega_\Lambda(a)$, the extra $\frac{12}{55}\epsilon$ will mean γ is no longer a constant to first order in $\Omega_\Lambda(a)$. This suggests that a constant γ is no longer a viable approximation for solving the perturbation equations. Rather than face the difficulty in generalising to a redshift dependent γ, we note that the constancy of γ can be maintained at the appropriate level of approximation by simply changing the approximate solution such that

$$g(a) = e^{\int_0^a (\,da'/a')\left[\Omega_m^\gamma(a') - 1 + b\right]}\,. \tag{8.19}$$

The new constant b means that

$$\gamma \approx -(f(a) - 1 + b)/\Omega_\Lambda(a)\,, \tag{8.20}$$

and we see that if b is chosen to offset the constant terms in $(f(a) - 1)$, then γ will once again be a constant. For the case of pure background energy exchange,

$$b = -\frac{12\epsilon}{55}\,, \tag{8.21}$$

and hence our first modification to the growth rate is

$$f(a) \simeq \Omega_m^{\gamma}(a) - \frac{12\epsilon}{55}\,, \tag{8.22}$$

where γ is subject to a minor perturbation

$$\gamma = \frac{6}{11} - \frac{6}{55}\epsilon\,. \tag{8.23}$$

8.3.2 Dilution

The energy exchange does not only affect the background dynamics; it also generates new terms in the perturbation equations, so (8.8) is no longer valid. By adding a homogeneous contribution to the background density, the fractional contrast, δ, is suppressed. In this particular parameterisation of energy exchange, the density and velocity perturbation equations are greatly simplified from the general case presented in [144]:

$$\delta' + \epsilon\mathcal{H}\delta + \theta - 3\Phi' = 0; \tag{8.24}$$

$$\theta' + \mathcal{H}\theta - k^2\Psi = 0. \tag{8.25}$$

The differential equation governing $f(a)$ now becomes

$$\begin{aligned}\frac{df}{d\ln a} + \frac{f}{2}\frac{d\ln H^2}{d\ln a} + f(f+2) + 2\epsilon \\ + \frac{\epsilon}{2}\frac{d\ln H^2}{d\ln a} - \frac{3}{2}\Omega_m(a) = 0\end{aligned} \tag{8.26}$$

and carrying these new terms through gives

$$f(a) \approx 1 - \frac{2}{5}\epsilon + \left[1 - \frac{1}{2}\Omega_\Lambda(a)\right](1-4\epsilon)I(a)\,. \tag{8.27}$$

The integral $I(a)$ has already been evaluated (8.16), and so we find:

$$f(a) \approx 1 + \left(\frac{6}{11} - \frac{16}{55}\epsilon\right)\Omega_\Lambda(a) - \epsilon\left(\frac{2}{5} + \frac{1}{55}\right)\,, \tag{8.28}$$

or

$$f(a) = \Omega_m^{\gamma}(a) - \frac{23}{55}\epsilon\,, \tag{8.29}$$

where

$$\gamma = \frac{6}{11} - \frac{16}{55}\epsilon\,. \tag{8.30}$$

8.3.3 Inertial Drag

In the previous subsection we implicitly assumed that there is a pure energy transfer, from the perspective of the dark matter rest frame. If however the energy transfer occurs in the CMB frame, the newly formed dark matter will not be instilled with the appropriate bulk motion, and a drag term appears in the perturbation equations.

The density and velocity perturbation equations are then of the form

$$\delta' + \epsilon\mathcal{H}\delta + \theta - 3\Phi' = 0, \tag{8.31}$$

$$\theta' + \mathcal{H}\,(1+\epsilon)\,\theta - k^2\Psi = 0. \tag{8.32}$$

The physical interpretation of the $\epsilon\mathcal{H}\delta$ term in (8.31) is one of dilution, as addressed in the previous subsection. A second effect which also slows structure growth is the extra drag term in (8.32), $\epsilon\mathcal{H}\theta$. This arises due to the appearance of stationary matter, reducing the mean flow rate.

If we define new perturbations $\bar{\delta} = \delta a^\epsilon$ and $\bar{\theta} = \theta a^\epsilon$, then substituting these rescaled variables into (8.31) and (8.32) leaves us with

$$\bar{\delta}' + \bar{\theta} = 0\,, \tag{8.33}$$

$$\bar{\theta}' + \mathcal{H}\bar{\theta} - k^2\bar{\Psi} = 0\,. \tag{8.34}$$

where we have neglected Φ', and the gravitational source term is rescaled as

$$-k^2\bar{\Psi} = \frac{3}{2}\mathcal{H}^2\Omega_m\bar{\delta} = \frac{3}{2}\mathcal{H}^2\Omega_m\delta a^\epsilon = -k^2\Psi a^\epsilon\,. \tag{8.35}$$

These two new equations, (8.33) and (8.34), are precisely the equations for perturbations as if no energy exchange were taking place beyond a background level. The previous section provided us with an approximate solution in a universe with background energy exchange:

$$\bar{\delta}' \simeq \bar{\delta}\mathcal{H}\left[\Omega_m(a)^\gamma - \epsilon\left(\frac{1}{5} + \frac{1}{55}\right)\right]\,. \tag{8.36}$$

After rescaling:

$$\delta' \simeq \delta\mathcal{H}\left[\Omega_m(a)^\gamma - \epsilon\left(\frac{6}{5} + \frac{1}{55}\right)\right]\,, \tag{8.37}$$

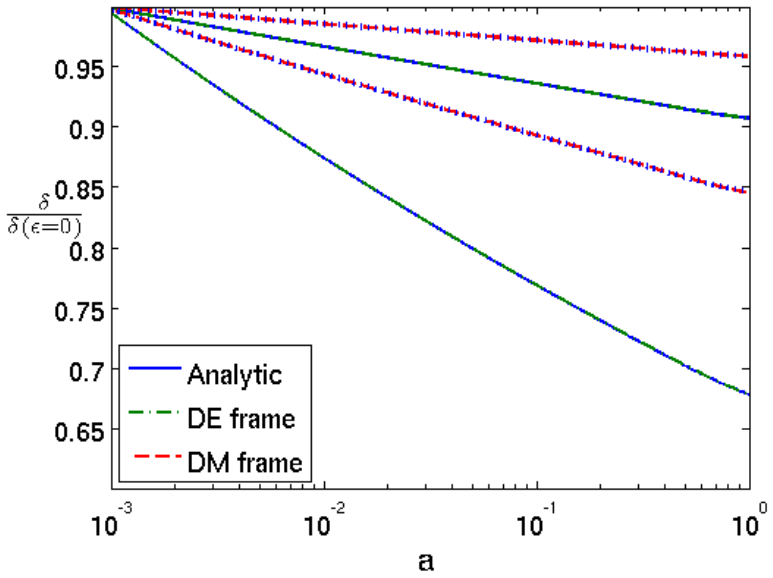

Figure 8.1: The fractional change in the evolution of linear perturbations in the presence of a decaying cosmological constant with $\epsilon = 0.01$ (upper) and $\epsilon = 0.04$ (lower), as defined in (8.4). The pair of dash-dotted lines highlight the slight suppression of growth induced by the diluting effect of transferring homogeneous energy into the dark matter frame. The dashed lines correspond to an energy transfer in the CMB rest frame, with the extra deceleration arising due to the introduction of stationary matter. These are well described by the solid lines, which illustrate the analytic solution given by (8.38).

or simply

$$f(a) = \Omega_m^\gamma(a) - \frac{67}{55}\epsilon\,, \tag{8.38}$$

and we return to the solution (8.23) to find

$$\gamma = \frac{6}{11} - \frac{6}{55}\epsilon\,. \tag{8.39}$$

We see from Figure 8.1 that this analytic approximation is very successful when compared to the numerical one.

The physical mechanisms contributing to this new growth rate are summarised in Table 8.1. The quoted total corresponds to the model where Q is unperturbed, and with the energy transfer occurring in the dark energy rest frame (that of the CMB). If instead the dark energy decay is sensitive to local fluctuations in the matter density, $\delta Q \propto \mathcal{H}\delta\rho_m$, the dilution term is no longer present. Similarly, if the energy exchange were to take place in the dark matter rest frame as opposed to the dark energy rest frame, the drag contribution vanishes.

Contribution	$\Delta f/\epsilon$	$\Delta\gamma/\epsilon$
Background	$-\left(\frac{1}{5}+\frac{1}{55}\right)$	$-\frac{6}{55}$
Dilution	$-\frac{1}{5}$	$-\frac{10}{55}$
Drag	$-\frac{4}{5}$	$+\frac{10}{55}$
Total	$-\frac{67}{55}$	$-\frac{6}{55}$

Table 8.1: Summary of the contributions to the growth rate from a decaying dark energy component, in units of ϵ. Whether the effects of dilution and drag arise will depend on the behaviour and orientation of the interaction parameter Q. Note that the terms in the left column, those contributing directly to $f(z)$, have a considerably greater impact than the corrections to γ in the right column.

8.3.4 Baryonic Correction

Thus far we have assumed the universe comprises solely of dark matter and dark energy, but before considering the observational consequences we must first extend our analysis to include baryons. We therefore divide the matter component into cold dark matter and baryonic terms, denoted ρ_c and ρ_b. The baryonic perturbations are simply given by

$$\delta_b'' + \mathcal{H}\delta_b' = 4\pi G a(\rho_c\delta_c + \rho_b\delta_b). \tag{8.40}$$

The analytic solution in (8.38) now requires a minor rescaling to compensate for the introduction of baryonic matter. A simple weighting of the decay parameter ϵ by the fraction of mass to which it applies is sufficient.

$$f = \Omega_m^\gamma - \frac{67}{55}\bar{\epsilon}, \tag{8.41}$$

$$\bar{\epsilon} \equiv \frac{\Omega_c}{\Omega_m}\epsilon. \tag{8.42}$$

In Figure 8.3, this formalism is seen to provide an excellent description for the combined matter perturbation, although we note that a more thorough treatment of the perturbations is required at very high redshifts ($z > 100$), where the radiation

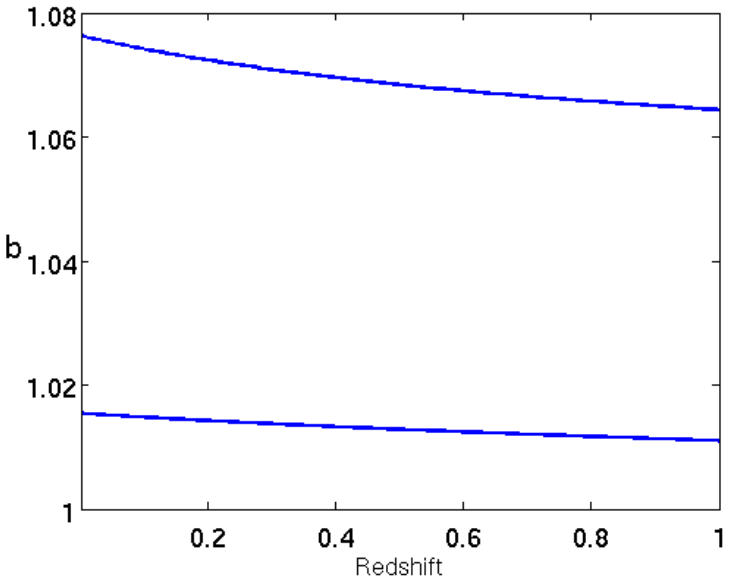

Figure 8.2: The baryonic bias induced by $\epsilon = 0.01$ and $\epsilon = 0.04$.

energy density starts to become significant. At later times the baryons closely track the growth rate of the dark matter perturbations, but with a slightly enhanced density contrast. Defining the baryonic bias $b \equiv \delta_b/\delta_c$, the present day value of this ratio is approximately given by

$$b \approx 1 + 2\epsilon, \tag{8.43}$$

as illustrated in Figure 8.2.

8.4 Observational consequences

We now review their impact on cosmological observables, using redshift-space distortions as a measure of the growth rate, before considering implications for the Integrated Sachs-Wolfe (ISW) effect.

8.4.1 Redshift space distortions

In order to highlight the observational consequences of this energy exchange, we evaluate the appropriate Fisher matrix for a redshift survey at $z = 0.5$ (see [132]), and combine this with the Dark Energy Task Force Fisher matrix for Planck. We marginalise over the parameter set

$$[w_0, w_a, \Omega_\Lambda, \Omega_k, \Omega_m h^2, \Omega_b h^2, n_s, A_s, \beta, \gamma, \sigma_p, \epsilon]. \tag{8.44}$$

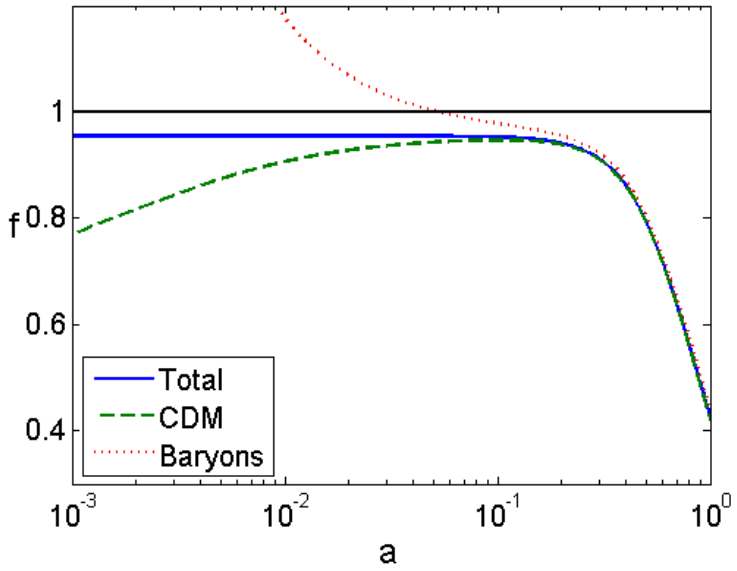

Figure 8.3: The logarithmic growth rate of both cold dark matter (dashed), and the baryons (dotted), in the presence of a decaying cosmological constant with $\epsilon = 0.04$. The solid line illustrates the approximate solution given by (8.41), which traces the total matter perturbation.

The standard cosmological parameters are taken to have fiducial values as derived from WMAP5 [54]. In order to remain consistent with the CMB Fisher matrix, when perturbing ϵ we ensure the value of ρ_m at $z = 1100$ is held fixed. The Hubble parameter h is slightly perturbed for the purposes of distance estimation. We neglect the small change induced by the Integrated Sachs-Wolfe effect (see 8.4.2).

As a parameterisation of modified gravity, the growth of structure is taken to follow the form given by [146, 100]

$$f \simeq \Omega_m^{\gamma}. \tag{8.45}$$

Clearly one ought to expect that, for the case of interacting models, applying the above prescription would lead to a biased estimate of the growth index γ. In addition to the extra ϵ term in (8.38), the value of γ is biased if the evolution of $\Omega_m(z)$ does not run as expected. This shift is illustrated by the solid contours in Figure 8.4 which are generated with $\epsilon = 0.04$.

In practice, we must deal with the Fingers of God in greater detail than a single parameter σ_p. It should also be noted that these models may have some impact on virialised structures, though we leave this as a topic for future investigation.

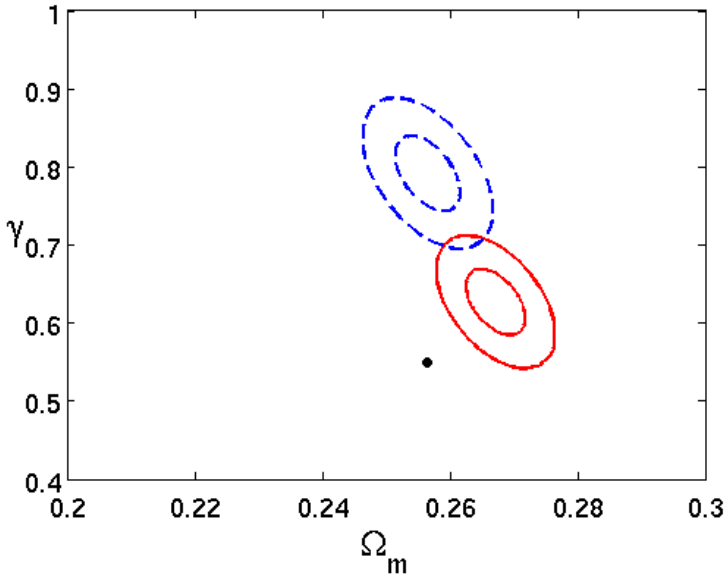

Figure 8.4: The solid contours provide an example of the bias which may be induced in the gravitational growth index γ when making the false assumption that dark energy is stable. The true model, as indicated by the black dot, corresponds to $\epsilon = 0.01$. The dashed contours demonstrate the modification to the growth index induced by the elastic interaction model outlined in [133].

8.4.2 Integrated Sachs-Wolfe effect

One restriction that dark energy models must satisfy is not to overpredict the Integrated Sachs-Wolfe effect. Excessive change in the gravitational potential would invariably generate large fluctuations in both the CMB and CMB-LSS cross-correlation on large angular scales. Indeed this has already been employed by Valiviita et al. [143] to assist in constraining models of decaying dark energy.

Since we are still working in the context of General Relativity, the gravitational potential is readily evaluated via the Poisson equation

$$\Phi(k,a) = -4\pi G\rho(t)a^2\frac{\delta(a)}{k^2}\,, \tag{8.46}$$

$$\Psi = \Phi\,. \tag{8.47}$$

This evolution is illustrated in Figure 8.5 for the standard case of flat ΛCDM, alongside small perturbations in the decay parameter (ϵ) and global curvature (Ω_k). While there is a distinctive change in behaviour at high redshift, this is a regime that lies out of

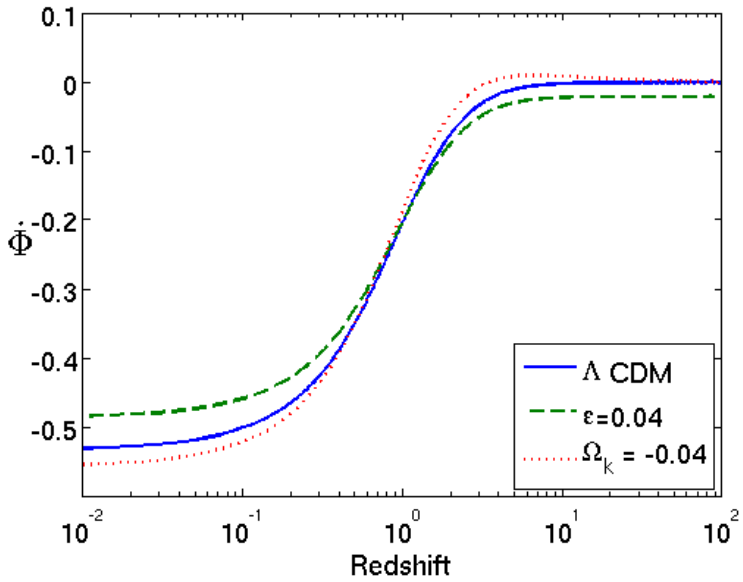

Figure 8.5: The decay rate of the gravitational potential as experienced in a flat ΛCDM Universe (solid), and with perturbed cosmologies $\epsilon = 0.01$ (dashed) and $\Omega_k = -0.01$ (dotted).

reach for cross-correlation studies. Radiation may also start to become significant at this point; we have not included this effect in our treatment.

The ISW signal is, at best, a $\sim 5\sigma$ observation, and as such there is significant room for flexibility in the anticipated signal strength. Taking a 20% change as the maximum permissible, this corresponds to an approximate upper limit of $\epsilon \lesssim 0.1$. This appears broadly consistent with the findings of Valiviita et al. [143], who established an upper bound of $|\Gamma| < 0.23H_0$ for a constant dark energy decay rate, from a combination of WMAP, supernovae, and BAO data. The two parameterisations are related by $\Gamma = -\mathcal{H}\epsilon$. Our earlier analysis on the growth rate may be naturally extended to this model, which we also find to be well described by the prescription

$$\Delta f \propto \Gamma/\mathcal{H}\,. \tag{8.48}$$

8.5 Discussion

In the case of a decaying cosmological 'constant', we have demonstrated that the growth rate of large scale structure is subject to a constant decrement $f = \Omega_m^\gamma - c$. For the simple model we consider, this can largely be attributed to the dragging effect on bulk motions induced by the production of stationary matter. Smaller contributions arise

from both the background dynamics, and the diluting effect of gradually introducing a homogeneous density field. It is also important to note that in more general interacting models the presence of dark energy perturbations may also influence the growth of structure. These act to enhance the growth rate, and may overpower the mechanisms considered here. Indeed in many cases the dark energy perturbations in coupled models lead to pathological instabilities.

Energy exchange within the dark sector leads to a change in the comoving matter density, which in turn induces a number of changes to cosmological observations. If these changes are not taken into consideration, then suppression to the growth of large scale structure leads to a naïvely inferred value of the growth index γ rising above the conventional value $\gamma > 0.55$. While the physical motivation for such models remains unclear, we believe this is no less true for current approaches to modified gravity.

As noted by Blandford et al. [24], a blind cosmologist living in the radiation dominated era might measure the evolution of the scale factor and erroneously conclude that the expansion is driven by a scalar field with an exponential potential. This thought experiment may be extended further: a "dark" astronomer only capable of studying the dark matter would notice a strange behaviour of non-linear structure. They might attempt to construct a modified theory of gravity which accounts for this behaviour. However the true source of this discrepancy is the momentum exchanged between the baryons and photons, inducing new features to the growth of structure. This modification is apparent both at the era of recombination, and the present day. We are at risk of falling into a similar trap here: the baryocentric assumption that neither dark energy nor dark matter exhibit any complex behaviour may lead to dark physics being mistaken for modified gravity.

Chapter 9

Conclusions and Summary

In this thesis we have explored some of the fundamental issues facing the standard model of cosmology, as well examined a variety of solutions and their implications.

In Chapter 4 we introduced an additional scalar field to the model proposed by Peebles and Vilenkin [115], using a symmetry breaking term to facilitate reheating in the universe. We constrained model parameters to a range, and calculated the reheating temperature as a function of these parameters. We also noted that the symmetry breaking causes the production, and eventual removal, of domain walls in the universe. An interesting avenue of further research would be to explore this aspect of the model; some preliminary numerical calculations we have carried out indicates these domain walls may persist as quasi-stable spatial deviations (solitons). A more rigorous calculation including the effects of gravity is still required to confirm this.

We moved on in Chapter 5 to address the effect of quantum corrections in models, such as the one described in Chapter 4, that include a coupling between the quintessence or inflation field, and another scalar field. There we described how to implement a mass-dependent renormalisation scheme in the context of cosmology, and noted that due to the large field values today for quintessence, and during during inflation, such quantum corrections may not be as large as when calculated in a mass-independent renormalisation scheme. We also highlighted the inherent problem of the missing physics in quintessence and inflation models, required to determine the renormalisation scale in the calculation of the effective potential.

Next we turned to cosmological perturbation theory, to try to understand the effect of energy exchange between dark matter and dark energy. We chose a simple form of energy exchange, and in Chapter 7 quantified the nature of an instability in this form of energy exchange. In Chapter 8 we examined a slightly different form of the energy exchange, and found that this effect becomes very difficult to distinguish observationally from modified gravity. We also provided a very successful analytic approximation to the growth rate of structure, that allowed us to break down effects of the interaction into separate parts, allowing for a clean physical interpretation of each case. Going beyond these simple forms for energy-momentum exchange is challenging, however, and leaves a great deal of room for further investigation. Carrying out the analysis with a form of energy exchange that is more suited to a scalar-field interaction would be particularly relevant to current lines of investigation [93].

Appendix A

Mass Dependent Renormalization Scheme

In this Appendix we summarize the mass dependent renormalization scheme [66] used to obtain the RGEs for mass parameters and couplings Eqs. (5.71)-(5.77). In order to set the notation and the procedure, we start by considering the simplest example, a single scalar field with potential:

$$V(\phi) = \frac{\lambda_\phi}{4!}\phi^4 + \frac{m_\phi^2}{2}\phi^2 \,. \tag{A.1}$$

The renormalized Lagrangian including the counterterms is then:

$$\begin{aligned} \mathcal{L}_{\rm ren} = \mathcal{L} + \delta\mathcal{L}_{ct} &= \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi) + \frac{\delta_{Z_\phi}}{2}\partial_\mu\phi\partial^\mu\phi - \frac{\delta m_\phi^2}{2}\phi^2 - \frac{\delta\lambda_\phi}{4!}\phi^4 - \delta\Omega \\ &= \frac{Z_\phi}{2}\partial_\mu\phi\partial^\mu\phi - \frac{Z^{-1}_{m_\phi^2}}{2}m_\phi^2\phi^2 - \frac{Z_\lambda}{4!}\lambda_\phi\phi^4 - Z_\Omega^{-1}\Omega \,, \end{aligned} \tag{A.2}$$

where ϕ, m_ϕ^2 and λ_ϕ are (finite) renormalized parameters, and $\phi_0 = Z_\phi^{1/2}\phi$, $m_{\phi 0}^2 = Z_\phi^{-1}Z^{-1}_{m_\phi^2}m_\phi^2$, $\lambda_{\phi 0} = Z_\phi^{-2}Z_{\lambda_\phi}\lambda_\phi$, the (infinite) bare parameters, and

$$\delta_{Z_\phi} = Z_\phi - 1 \,, \tag{A.3}$$

Z_ϕ being the wave function renormalization constant.

Loop corrections are computed with the Lagrangian given by the first line in Eq. (A.2), including the corrections given by the counterterms. The first step is then

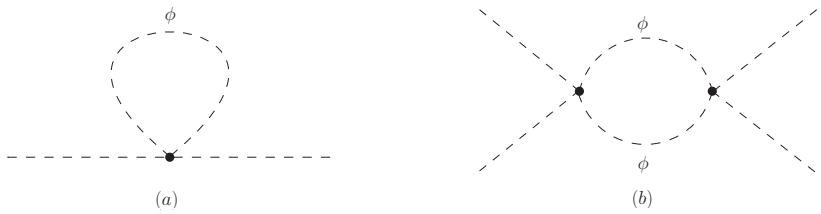

Figure A.1: (a) 1-loop scalar self-energy diagram. (b) 1-loop correction to the proper scalar quartic vertex.

to regularize the divergent integrals, and for that we use dimensional regularization: evaluating the integrals in $d = 4 - \epsilon$ dimensions, and then taking the limit ϵ going to zero. The divergent term at 1-loop is isolated as the single pole when $d = 4$ ($2/\epsilon$ term). The renormalized (finite) mass and coupling are defined by imposing suitable normalization conditions on the n-point 1PI Green functions $\Gamma^{(n)}$ at some arbitrary scale μ [122]. The relation between the bare and renormalized n-point 1PI functions is given by:

$$\Gamma^{(n)}(p^2) \mid_{p^2=\mu^2} = Z_\phi^{n/2} \Gamma_0^{(n)}(p^2) \mid_{p^2=\mu^2} , \tag{A.4}$$

where as before the subscript "0" denotes the bare quantity. The normalization condition fixes the counterterms that cancel out the divergent terms; this is equivalent to define the renormalization constants, given the relation between these and the counterterms introduced in section 5.3, Eqs.(5.65)-(5.67). We now derive them explicitly in the MDR scheme.

The 2-point 1PI renormalized function $\Gamma^{(2)}$, including the contributions from the counterterms and the radiative correction $\Pi^{(\phi)}(p^2)$ (Fig. (A.1.a)), is given by:

$$\Gamma^{(2)}(p^2) = p^2 - m_\phi^2 + \delta_{Z_\phi} p^2 - \delta m_\phi^2 + \Pi^{(\phi)}(p^2) . \tag{A.5}$$

The counterterms, or equivalently the renormalization constants, are fixed by demanding $\Gamma^{(2)}$ to be that of a free-field theory with running mass parameter $m_\phi^2(\mu)$ [139, 7], at the renormalization scale μ:

$$\Gamma^{(2)}(p^2) \mid_{p^2=\mu^2} \equiv (p^2 - m_\phi^2(p^2)) \mid_{p^2=\mu^2} , \tag{A.6}$$

where p^2 is the incoming Euclidean momentum[1]. The 1-loop contribution is given in

[1] In this Appendix p^2 will refer to the Euclidean momentum, and we drop for simplicity the subindex "E" hereon.

this case by:

$$\Pi^{(\phi)}(p^2) = \frac{\lambda_\phi}{2} L(m_\phi^2/\hat{\mu}^2) m_\phi^2 \,, \tag{A.7}$$

$$L(m_\phi^2/\hat{\mu}^2) = \frac{1}{16\pi^2}\left(\frac{2}{\bar{\epsilon}} + 1 - \ln\frac{m_\phi^2}{\hat{\mu}^2}\right) , \tag{A.8}$$

where the scale $\hat{\mu}$ is introduced in the regularization procedure because of dimensional reasons, and $2/\bar{\epsilon} = 2/\epsilon - \gamma_E + \ln 4\pi$. When the scalar field does not couple to fermions, the wave function renormalization constant at 1-loop does not receive any contribution and therefore:

$$Z_\phi = 1 \,. \tag{A.9}$$

On the other hand, the normalization condition fixes the mass counterterm:

$$\delta m_\phi^2 = \Pi^{(\phi)}(\mu^2) \,, \tag{A.10}$$

and from Eq. (5.65) we can read the mass renormalization constant:

$$Z_{m_\phi^2} = 1 - \frac{\lambda_\phi}{2} L(m_\phi^2/\hat{\mu}^2) \,. \tag{A.11}$$

The relation between the renormalized mass parameter and the bare parameter is given by:

$$m_\phi^2(\mu) = Z_\phi Z_{m_\phi^2} m_{\phi 0}^2 \,, \tag{A.12}$$

and thus, taking the derivative with respect to the renormalization scale, one obtains the RGE for the mass parameter:

$$\beta_{m_\phi^2} = \frac{dm_\phi^2(\mu)}{d\ln\mu} = m_\phi^2 \frac{d\ln Z_{m_\phi^2}}{d\ln\mu} = 0 \,. \tag{A.13}$$

Thus, given that at 1-loop the radiative correction $\Pi^{(\phi)}(p^2)$ is independent of the external momentum, this (quadratically) divergent contribution can be reabsorbed into a redefinition of the mass parameter.

For the renormalized vacuum energy, the situation is quite similar. The zero-point 1PI function at 1-loop are given by vacuum diagrams which do not depend on any scale except that of the mass of the particle in the loop. Therefore,

$$Z_\Omega \Omega = \Omega - \frac{m_\phi^4}{4} L(m_\phi^2/\hat{\mu}^2) \,, \tag{A.14}$$

and

$$\beta_\Omega = 0 \,. \tag{A.15}$$

The vacuum contribution will be fixed by the boundary conditions, say $\Omega(\mu) = 0$. Either at higher orders, or when the scalar couples to fermions, we will have $Z_\phi \neq 1$ and $\beta_{m_\phi^2} \neq 0$, but still the pure vacuum contributions give $\beta_\Omega = 0$.

The RGE for the quartic coupling is derived in a similar manner, obtaining the coupling renormalization constant by imposing the normalization condition on the 4-point 1PI function. The renormalized 4-point 1PI function at 1-loop is given by:

$$\Gamma^{(4)}(p^2) = -\lambda_\phi - \delta\lambda_\phi + \frac{3}{2}\lambda_\phi^2 \Gamma(p^2, m_\phi^2) \tag{A.16}$$

where $\Gamma(p^2, m^2)$ is the contribution from the 1-loop diagram Fig. (A.1.b):

$$\Gamma(p^2, m^2) = \frac{1}{16\pi^2}\left(\frac{2}{\bar{\epsilon}} - \int_0^1 dx \ln \frac{m^2 + x(1-x)p^2}{\hat{\mu}^2}\right). \tag{A.17}$$

We impose the normalization condition:

$$\Gamma^{(4)}(p^2) \mid_{p^2=\mu^2} \equiv -\lambda_\phi(\mu)\,, \tag{A.18}$$

which defines the coupling counterterm:

$$\delta\lambda_\phi = \frac{3}{2}\lambda_\phi^2 \Gamma(\mu^2, m_\phi^2)\,, \tag{A.19}$$

and using Eq. (5.66), the coupling renormalization constant $Z_{\lambda_\phi}^{-1}$:

$$Z_{\lambda_\phi}^{-1} = 1 - \frac{3}{2}\lambda_\phi \Gamma(\mu^2, m_\phi^2)\,. \tag{A.20}$$

The relation between renormalized and bare coupling is given by:

$$\lambda_\phi(\mu) = Z_\phi^2 Z_{\lambda_\phi}^{-1} \lambda_{\phi 0}\,, \tag{A.21}$$

and as before, taking the derivative the RGE for the coupling reads:

$$\frac{d\lambda_\phi(\mu)}{d\ln\mu} = \frac{3\lambda_\phi^2}{(4\pi)^2} F_2(a_\phi)\,, \tag{A.22}$$

$$F_2(a_\phi) = 1 - \frac{2a_\phi}{\sqrt{1+4a_\phi}} \ln \frac{\sqrt{1+4a_\phi}+1}{\sqrt{1+4a_\phi}-1}\,, \tag{A.23}$$

where $a_\phi = m_\phi^2/\mu^2$. Threshold effects are included in the effective coupling through the momentum dependence $p^2 = \mu^2$ of the radiative corrections[2], such that the coefficients of the RGEs are modulated by a threshold function $F_2(a)$. The latter reduces to one in the massless limit, $a = 0$, but it goes to zero when $a = m^2/\mu^2$ goes to infinity. That is,

[2]In practice, the running effective coupling can be obtained by taking the derivative of the 1-loop 1PI Green functions with respect to the momentum, and then replacing $p^2 = \mu^2$.

in the massless limit one recover the same RGEs for the effective couplings than those computed in a mass independent scheme, like the $\overline{\rm MS}$ scheme. But in the opposite limit, for a heavy state with $a \gg 1$, the contribution is more and more suppressed as the ratio a increases.

Having set the scheme in the simplest model, we can extend it now to the case of study, adding the second scalar field χ, with potential:

$$V(\phi,\chi) = \Omega + \frac{1}{2}m_\phi^2\phi^2 + \frac{\lambda_\phi}{4!}\phi^4 + \frac{1}{2}m_\chi^2\chi^2 + \frac{\lambda_\chi}{4!}\chi^4 + \frac{1}{2}g^2\chi^2\phi^2\,. \tag{A.24}$$

Adding the corresponding counterterms, the renormalized Lagrangian for the scalar sector is then:

$$\begin{aligned}\mathcal{L}_S &= \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi - V(\phi,\chi)\\ &+\frac{\delta_{Z_\phi}}{2}\partial_\mu\phi\partial^\mu\phi + \frac{\delta_{Z_\chi}}{2}\partial_\mu\chi\partial^\mu\chi - \frac{\delta m_\phi^2}{2}\phi^2 - \frac{\delta m_\chi^2}{2}\chi^2 - \frac{\delta\lambda_\phi}{4!}\phi^4 - \frac{\delta\lambda_\chi}{4!}\chi^4\\ &-\frac{\delta g^2}{2}\phi^2\chi^2 - \delta\Omega\,.\end{aligned} \tag{A.25}$$

We also consider Yukawa interactions between χ and N_F fermions Ψ_α, with the fermionic Lagrangian given by:

$$\begin{aligned}\mathcal{L}_F &= \bar\Psi_\alpha(i\gamma^\mu\partial_\mu - m_f)\Psi_\alpha - h\chi\bar\Psi_\alpha\Psi + \delta_{Z_f}\bar\Psi_\alpha\Psi\\ &-\delta m_f\bar\Psi_\alpha\Psi - \delta h\chi\bar\Psi_\alpha\Psi\,,\end{aligned} \tag{A.26}$$

where we have taken for simplicity a common mass m_f and Yukawa coupling h for the fermions.

The interaction term given by the coupling g^2 only adds a constant (quadratically) divergent term to Π^ϕ, through the same diagram than in Fig. (A.1.a) but with a χ internal line. Therefore, still we have $Z_\phi = 1$ and

$$m_\phi^2 Z_{m_\phi^2} = m_\phi^2 - \frac{\lambda_\phi}{2}m_\phi^2 L(m_\phi^2/\hat\mu^2) - g^2 m_\chi^2 L(m_\chi^2/\hat\mu^2)\,, \tag{A.27}$$

and thus $\beta_{m_\phi^2} = 0$. On the other hand, Z_χ and $Z_{m_\chi^2}$ receive a p-dependent contribution from the loop of fermions. The diagrams contributing to the χ field 2-point function are given in Fig. (A.2), with

$$\begin{aligned}\Pi^{(\chi)}(p^2) &= 2h^2N_F\Gamma(p^2,m_f^2)p^2 + 8h^2N_F\Gamma(p^2,m_f^2)m_f^2 + \frac{\lambda_\chi}{2}L(m_\chi^2/\hat\mu^2)m_\chi^2\\ &+g^2L(m_\phi^2/\hat\mu^2)m_\phi^2\,.\end{aligned} \tag{A.28}$$

Figure A.2: 1-loop self-energy diagram contribution to the χ propagator. Dashed lines represents scalars, and fermions are given by solid lines.

For the couplings, the renormalization constants Z_χ and $Z_{m^2_\chi}$ are given then by:

$$Z_\chi = 1 - 2h^2 N_F \Gamma(\mu^2, m_f^2)\,, \tag{A.29}$$

$$\begin{aligned} m^2_\chi Z_{m^2_\chi} &= m^2_\chi - 8h^2 N_F m_f^2 \Gamma(\mu^2, m_f^2) - \frac{\lambda_\chi}{2} m^2_\chi L(m^2_\chi/\hat{\mu}^2) \\ &\quad - g^2 m^2_\phi L(m^2_\phi/\hat{\mu}^2)\,. \end{aligned} \tag{A.30}$$

The renormalization constant $Z^{-1}_{\lambda_\phi}$ picks up a new term similar to that in Eq. (A.20) due to the loop with 2 massive χ states, i.e., a diagram similar to that in Fig. (A.1.b) but with χ running in the loop. The 1-loop diagrams contributing to the λ_χ and g^2 coupling renormalization constants are given in Figs. (A.3.a) and (A.3.b) respectively. The set of renormalization constants are then:

$$\lambda_\phi Z^{-1}_{\lambda_\phi} = \lambda_\phi - \frac{3}{2}\lambda^2_\phi \Gamma(\mu^2, m^2_\phi) - 6g^4 \Gamma(\mu^2, m^2_\chi)\,, \tag{A.31}$$

$$\lambda_\chi Z^{-1}_{\lambda_\chi} = \lambda_\chi - \frac{3}{2}\lambda^2_\chi \Gamma(\mu^2, m^2_\chi) - 6g^4 \Gamma(\mu^2, m^2_\phi) - 24h^4 N_F \Gamma(\mu^2, m_f^2)\,, \tag{A.32}$$

$$g^2 Z^{-1}_{g^2} = g^2 \left(1 - \frac{\lambda_\chi}{2}\Gamma(\mu^2, m^2_\chi) - \frac{\lambda_\phi}{2}\Gamma(\mu^2, m^2_\phi) - 4g^2 \Gamma_2(\mu^2, m^2_\chi, m^2_\phi)\right)\,, \tag{A.33}$$

where:

$$\Gamma_2(p^2, m_1^2, m_2^2) = \frac{1}{16\pi^2}\left(\frac{2}{\bar{\epsilon}} - \int_0^1 dx \ln \frac{m_1^2 x + m_2^2(1-x) + x(1-x)p^2}{\hat{\mu}^2}\right)\,. \tag{A.34}$$

To obtain the RGE for the Yukawa coupling we also need to renormalize the inverse fermion propagator S^{-1} and the Yukawa vertex $\Gamma^{(3)}$. The corresponding 1-loop diagrams are given in Figs. (A.4.a) and (A.4.b). The normalization condition for S^{-1} and $\Gamma^{(3)}$ are similar to Eqs. (A.6) and (A.18):

$$S^{-1}(p^2)_{p^2=\mu^2} \equiv (\not{p} - m_f(p^2))\mid_{p^2=\mu^2}\,, \tag{A.35}$$

$$\Gamma^{(3)}(p^2)_{p^2=\mu^2} \equiv h_R(\mu)\,, \tag{A.36}$$

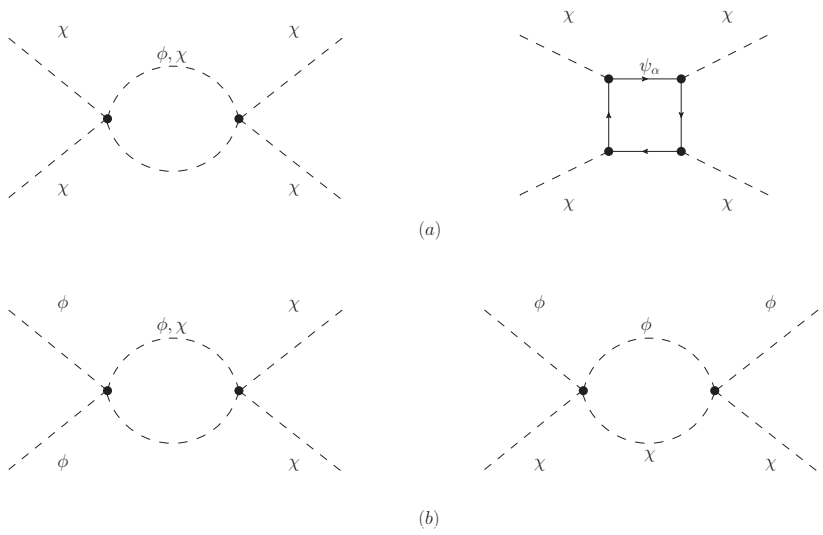

Figure A.3: (a) 1-loop proper vertex correction to λ_χ. (b) 1-loop proper vertex correction to g^2.

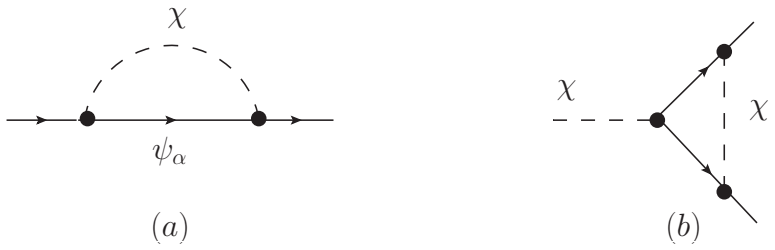

Figure A.4: (a) 1-loop fermion self-energy. (b) 1-loop Yukawa vertex correction.

and from the above equations we can obtain Z_f, Z_h, and Z_{m_f} for massive fermions:

$$Z_f = 1 - \frac{h^2}{2}\Gamma_f(\mu^2, m_\chi^2, m_f^2)\,, \tag{A.37}$$

$$Z_h^{-1} = 1 - h^2\Gamma_2(\mu^2, m_\chi^2, m_f^2)\,, \tag{A.38}$$

$$Z_{m_f} = 1 - h^2\Gamma_f(\mu^2, m_\chi^2, m_f^2)\,, \tag{A.39}$$

where:

$$\Gamma_f(p^2, m_1^2, m_2^2) = \frac{1}{16\pi^2}\left(\frac{2}{\bar{\epsilon}} - \int_0^1 dx x \ln\frac{m_1^2 x + m_2^2(1-x) + x(1-x)p^2}{\hat{\mu}^2}\right) \tag{A.40}$$

Finally, from the above renormalization constant, and the relations:

$$
\begin{aligned}
m_\phi^2(\mu) &= Z_\phi Z_{m_\phi^2} m_{\phi 0}^2 \,, && \text{(A.41)}\\
m_\chi^2(\mu) &= Z_\chi Z_{m_\chi^2} m_{\chi 0}^2 \,, && \text{(A.42)}\\
\lambda_\phi(\mu) &= Z_\phi^2 Z_{\lambda_\phi}^{-1} \lambda_{\phi 0} \,, && \text{(A.43)}\\
\lambda_\chi(\mu) &= Z_\chi^2 Z_{\lambda_\chi}^{-1} \lambda_{\chi 0} \,, && \text{(A.44)}\\
g^2(\mu) &= Z_\phi Z_\chi Z_{g^2}^{-1} g_0^2 \,, && \text{(A.45)}\\
h^2(\mu) &= Z_\chi^{1/2} Z_f Z_{h^2}^{-1} h_0^2 \,, && \text{(A.46)}
\end{aligned}
$$

the RGEs (5.71)-(5.76) including the threshold functions are easily derived (setting $m_f = 0$). The latter are given by:

$$
\begin{aligned}
F_1(a) &= -(4\pi)^2 \frac{d\Gamma_2(p^2, m^2, 0)}{d \ln p^2} = 1 - a \ln \frac{1+a}{a} \,, && \text{(A.47)}\\
F_2(a) &= -(4\pi)^2 \frac{d\Gamma(p^2, m^2)}{d \ln p^2} = 1 - \frac{2a}{\sqrt{1+4a}} \ln \frac{\sqrt{1+4a}+1}{\sqrt{1+4a}-1} \,, && \text{(A.48)}\\
F_3(a) &= -(4\pi)^2 \left(\frac{d\Gamma_f(p^2, m^2, 0)}{d \ln p^2} + 2 \frac{d\Gamma_2(p^2, m^2, 0)}{d \ln p^2} \right)\\
&= 1 + 2a(1 - (1+a) \ln \frac{1+a}{a}) \,, && \text{(A.49)}
\end{aligned}
$$

where in computing $F_1(a)$ from the loop with one light m_ϕ and a heavy $m_\chi \gg m_\phi$ we have set $m_\phi = 0$. $F_2(a)$ is the threshold function for a scalar loop with 2 equal massive states, $F_1(a)$ that of a scalar loop with one massless and one massive scalar state, and $F_3(a)$ that with massless fermions and one massive scalar.

Appendix B

Power Suppression of 2-loop Coefficients

At 2-loop order, the RG-improved effective potential is given by the 1-loop effective potential with running parameters evaluated using the 2-loop RG equations. Following the MDR prescription, again the optimal choice to fix the renormalization scale is below all massive thresholds; thus the 2-loop effective potential reduces to the tree-level potential plus the 2-loop RGE functions. Decoupling will be included in the latter through threshold functions, similarly to the 1-loop RGEs.

At 2-loop order, both the wave function renormalization constant Z_ϕ and the mass parameter one $Z_{m_\phi^2}$ get a μ dependent contribution from the sunset diagram in Fig. (B.1.a), whereas the quartic coupling λ_ϕ correction comes from Fig. (B.1.b). The 2-loop beta functions in a mass independent scheme can be found for example in Refs. [44, 105, 106, 62, 61], where one only would need to extract the divergent contributions from those diagrams. In the MDR scheme we need to carry out the full calculation of the diagram keeping the finite contributions. We will not attempt such a full 2-loop calculation here, and we only want to argue that such diagrams gives a power suppression of the corresponding RGE coefficients that go at least like $O(\mu^2/M_\alpha^2)$ when $M_\alpha^2 \gg \mu^2$, M_α^2 being the heavy mass running in the loop. For the first diagram in the vertex correction in Fig.(B.1.b), this can be viewed as the 1-loop vertex correction but with the LHS interaction replaced by an effective (momentum dependent) vertex

again like that of Fig. (A.1.b). According to the 1-loop calculation, this will give the corresponding suppression by the mass running in the loop also in the 2-loop coefficients. The contributions of the order of $O(\lambda_\phi g^4)$, and $O(g^6)$, will be therefore suppressed by the heavy mass M_χ^2. That of the order of $O(\lambda_\phi^3)$ is suppressed by factors $O(\mu^2/M_\phi^2)$ when $\mu \ll M_\phi$. In the cosmology models studied in the text of this paper, generally $M_\phi \sim \mu$ and so these diagrams are not suppressed due to decoupling effects, but rather because λ_ϕ is always tiny. The second diagram in Fig. (B.1.b) gives a term ($O(h^2 g^4)$ coming from the insertion of the fermion loop in one of the internal χ lines. Again, this contribution will be suppressed by a factor $O(\mu^2/M_\chi^2)$, similarly to the 1-loop vertex correction without the fermion insertion.

We will therefore concentrate on the sunset diagram in Fig. (B.1.a), and in particular on the contribution to the wave function renormalization constant. In general for the sunset diagram with three different internal scalar masses we have:

$$\Pi_S^\phi(p^2, m_1^2, m_2^2, m_3^2) = \frac{\lambda_i^2}{6}\left(\sum_{j=1,2,3} m_j^2 A_j(p^2, m_i^2) + p^2 B(p^2, m_i^2)\right), \tag{B.1}$$

$$(4\pi)^4 A_j(p^2, m_i^2) = -2\left(\frac{1}{\epsilon^2} + \frac{1}{\epsilon}\left(\frac{3}{2} - \ln\frac{m_j^2}{4\pi\hat{\mu}^2} - \gamma_E + I_m(p^2, m_i^2)\right)\right), \tag{B.2}$$

$$(4\pi)^4 B(p^2, m_i^2) = \frac{1}{2}\Big(\frac{1}{\epsilon} - \gamma_E + \frac{1}{2} - 2\int_0^1 dx \int_0^1 dy (1-y) \ln\left(M^2(x,y) + p^2 y(1-y)\right)\Big), \tag{B.3}$$

where $M^2(x,y) = (m_1^2 x + m_2^2(1-x))y/(x(1-x)) + m_3^2(1-y)$, and λ_i is a general quartic coupling with the appropriate symmetry factors. In particular, when we have three ϕ running in the loop then $\lambda_i^2 = \lambda_\phi^2$, and with two χ and one ϕ then $\lambda_i^2 = 12 g^4$. The function $I_m(p^2, m_i^2)$ is the finite contribution to the mass renormalization. By adding the sunset contribution to the 2-point function Eq. (A.5), the normalization condition Eq. (A.6) fixes the non-vanishing 2-loop wave function counterterm $\delta_Z^{(2)}$, which defines the wave function renormalization constant $Z_\phi^{(2)} = 1 + \delta_Z^{(2)}$, and then the anomalous dimension of the field:

$$\gamma_\phi^{(2)} = \frac{d\ln Z_\phi^{(2)}}{d\ln\mu^2} = \frac{1}{(4\pi)^4}\left(\frac{\lambda_\phi^2}{12} F_{22}(a_\phi, a_\phi, a_\phi) + g^4 F_{22}(a_\chi, a_\chi, a_\phi)\right), \tag{B.4}$$

with the threshold function given by the parametric integral:

$$F_{22}(a_1, a_2, a_3) = 2\int_0^1 dx \int_0^1 dy \frac{y(1-y)^2}{a(x,y) + y(1-y)}, \tag{B.5}$$

where $a(x,y) = M^2(x,y)/\mu^2$. From the above expression, one can check that indeed when we have only massless states running in the loop then $F_{22}(0) = 1$, but when any

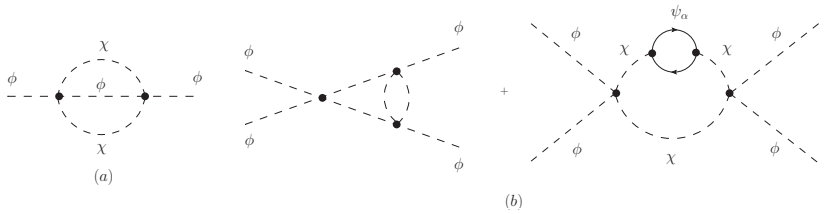

Figure B.1: (a) 2-loop scalar wave function renormalization diagram.(b) 2-loop correction to the proper scalar quartic vertex.

of the masses a_i goes to infinity, then $F_{22}(a_i)$ goes to zero as $O(1/a_i)$. The integral can be evaluated numerically for arbitrary mass parameters, and it behaves similarly to the 1-loop threshold functions in Fig. (5.1).

From the renormalization condition Eq. (A.43), the 2-loop beta function for λ_ϕ gets contributions from the anomalous dimension $\gamma_\phi^{(2)}$ and the proper vertex $\gamma_V^{(2)} = -d\ln Z_{\lambda_\phi}^{(2)}/d\ln\mu$,

$$\begin{aligned}
\beta_{\lambda_\phi}^{(2)} &= 4\lambda_\phi\gamma_\phi^{(2)} + \gamma_V^{(2)} \\
&= \frac{1}{(4\pi)^4}\left(\lambda_\phi^3\left(\frac{1}{3}F_{22}(a_\phi,a_\phi,a_\phi) - 6G(a_\phi,a_\phi)\right)\right. \\
&\quad +\lambda_\phi g^4(4F_{22}(a_\chi,a_\chi,a_\phi) - 24G(a_\phi,a_\chi) - 96g^6G(a_\chi,a_\chi) \\
&\quad \left. -432N_F h^2 g^4 G_F(a_f,a_\chi)\right).
\end{aligned} \tag{B.6}$$

We have not computed explicitly the 2-loop vertex threshold functions $G(a_i,a_j)$, $G_F(a_i,a_h)$, but as argued above, we can expect them to have the correct limits when $a_i \gg 1$, and therefore decoupling in the sense of power suppression is also maintained in the MDR scheme at the 2-loop level.

Appendix C

1-loop Effective Potential Within the Mass Dependent Renormalization (MDR) Scheme

The 1-loop radiative correction to the effective potential, computed using dimensional regularization, is given by:

$$\Delta V^{(1)}_{\rm reg} = -\frac{1}{64\pi^2}\left(M_\phi^4(L(M_\phi^2/\hat{\mu}^2)+\frac{1}{2}) + M_\chi^4(L(M_\chi^2/\hat{\mu}^2)+\frac{1}{2})\right). \tag{C.1}$$

The renormalized 1-loop effective potential is obtained by adding to $\Delta V^{(1)}_{\rm reg}$ the appropriate counterterms:

$$\Delta V^{(1)} = \Delta V^{(1)}_{\rm reg} + (Z_\Omega^{-1}-1)\Omega + \frac{1}{2}m_\phi^2(Z_{m_\phi^2}^{-1}-1)\phi^2 + \frac{\lambda_\phi}{4!}(Z_{\lambda_\phi}-1)\phi^4\,. \tag{C.2}$$

By plugging $Z_\phi = 1$, computing the renormalization constants given in Eqs. (A.27), (A.31) in the MDR scheme, and renormalizing the cosmological constant by demanding $V(\phi=0)=0$, one has:

$$\begin{aligned}\Delta V^{(1)} &= \frac{1}{64\pi^2}\left(m_\phi^4\ln\frac{M_\phi^2}{m_\phi^2} + \lambda_\phi m_\phi^2\phi^2\ln\frac{M_\phi^2}{m_\phi^2} + m_\chi^4\ln\frac{M_\chi^2}{m_\chi^2} + 2g^2m_\chi^2\phi^2\ln\frac{M_\chi^2}{m_\chi^2}\right.\\ &\left. + \frac{\lambda_\phi^2\phi^4}{4}\left(\ln\frac{M_\phi^2}{\mu^2} - I(m_\phi^2/\mu^2)\right) + g^4\phi^4\left(\ln\frac{M_\chi^2}{\mu^2} - I(m_\chi^2/\mu^2)\right)\right),\end{aligned} \tag{C.3}$$

where

$$I(a) = \int_0^1 dx \ln(a + x(1-x)) = \ln a - 2 - \sqrt{1+4a} \ln \frac{\sqrt{1+4a}-1}{\sqrt{1+4a}+1} \,. \tag{C.4}$$

Because the 1-loop renormalization constant have been computed in the previous section in the symmetric phase of the theory (i.e., taking the vev of the field to vanish), the threshold functions $I(a)$ depend on the mass parameters m_i^2, instead of the physical effective masses M_i^2 relevant for the effective potential. Then, this expression would only lead to decoupling of heavy states when $m_\phi \,, m_\chi \gg \phi$, but not when $\phi \gg m_\phi$, m_χ. For example by taking $\mu \ll m_\phi$, m_χ, and expanding the threshold function $I(a)$ when $a \gg 1$, we have:

$$\begin{aligned} \Delta V^{(1)} &= \frac{1}{64\pi^2} \left(M_\phi^4 \ln \frac{M_\phi^2}{m_\phi^2} + M_\chi^4 \ln \frac{M_\chi^2}{m_\chi^2} \right. \\ &\left. + \frac{\lambda_\phi^2 \phi^4}{4} \left(-\frac{\mu^2}{6m_\phi^2} + \cdots \right) + g^4 \phi^4 \left(-\frac{\mu^2}{6m_\chi^2} + \cdots \right) \right) \,, \end{aligned} \tag{C.5}$$

and unless $m_\phi \sim m_\chi$, we are still left with potentially large logs, $\ln \phi^2/m_\phi^2$, $\ln \phi^2/m_\chi^2$, and the original problem addressed in Refs. [16, 83, 82]. On the other hand, in the particular limit that the mass parameters vanish, $m_\phi = m_\chi = 0$, one just recover the expression for the effective potential computed in a mass independent scheme:

$$\Delta V^{(1)} = \frac{1}{64\pi^2} \left(\frac{\lambda_\phi^2 \phi^4}{4} \left(\ln \frac{\lambda_\phi \phi^2/2}{\mu^2} - 2 \right) + g^4 \phi^4 \left(\ln \frac{g^2 \phi^2}{\mu^2} - 2 \right) \right) \,. \tag{C.6}$$

In this case, given that both mass scales are set by the vev of the field ϕ, large logs could be controlled and resummed by taking for example $\mu \simeq \phi$.

This apparent failure of the MDR scheme can be related to the fact that the effective potential is computed in the non-symmetric phase of the theory, after shifting the field ϕ by its vev. One should therefore also impose the renormalization conditions and get the counterterms in this phase. After shifting the field, the propagators running in the loops depend now on the effective mass and, by repeating the calculation done in the previous section, one can derive similarly the renormalization constants now with the threshold functions depending on M_α. This accounts to replace m_i^2 by M_i^2 in both the logs and threshold functions in Eq. (C.3), and the effective potential is then given by:

$$\Delta V^{(1)} = \frac{1}{64\pi^2} \left(\frac{\lambda_\phi^2 \phi^4}{4} (\ln \frac{M_\phi^2}{\mu^2} - I(M_\phi^2/\mu^2)) + g^4 \phi^4 (\ln \frac{M_\chi^2}{\mu^2} - I(M_\chi^2/\mu^2)) \right) \tag{C.7}$$

This is similar to the effective potential obtained in Ref. [34]. In that work, decoupling is introduced in the effective potential through step-functions at each physical threshold

$M_i = \mu$, Eq. (5.63). In our case it is implemented through the threshold function $I(a_i)$ obtained when computing the 1-loop radiative corrections.

Finally, the RG-improved effective potential is given by absorbing the log dependence on the running parameters, such that the 1-loop potential is given by the tree-level potential evaluated at the boundary $t_* = \ln \mu_*/\mu$:

$$V^{\text{eff}} = \frac{1}{2} m_\phi^2(t_*)\phi^2(t_*) + \frac{\lambda_\phi(t_*)}{4!}\phi^4(t_*)\,. \tag{C.8}$$

This can be shown by explicit integration of the RGEs, plugging the result in the tree-level potential. At 1-loop order, the mass parameter and the field do not run, $m_\phi(t_*) = m_\phi$, and $\phi(t_*) = \phi$, and integrating the RGE for λ_ϕ in the MDR scheme, Eq. (5.71) we have:

$$\begin{aligned}\lambda_\phi(\mu_*) &\simeq \lambda_\phi(\mu) + \frac{3\lambda_\phi^2(\mu)}{32\pi^2}\left(I(M_\phi^2/\mu_*^2) - I(M_\phi^2/\mu^2) - \ln\frac{\mu_*^2}{\mu^2}\right) \\ &\quad + \frac{12 g^4(\mu)}{32\pi^2}\left(I(M_\chi^2/\mu_*^2) - I(M_\chi^2/\mu^2) - \ln\frac{\mu_*^2}{\mu^2}\right)\,.\end{aligned} \tag{C.9}$$

Taking $\mu_* \ll M_\phi$, M_χ, this gives:

$$\lambda_\phi(\mu_*) \simeq \lambda_\phi(\mu) + \frac{3\lambda_\phi^2(\mu)}{32\pi^2}\left(\ln\frac{M_\phi^2}{\mu^2} - I(M_\phi^2/\mu^2)\right) + \frac{12 g^4(\mu)}{32\pi^2}\left(\ln\frac{M_\chi^2}{\mu^2} - I(M_\chi^2/\mu^2)\right)\,, \tag{C.10}$$

and therefore,

$$\begin{aligned} V^{\text{eff}} &= \frac{1}{2} m_\phi^2(t_*)\phi^2(t_*) + \frac{\lambda_\phi(t_*)}{4!}\phi^4(t_*) \\ &\simeq \frac{1}{2} m_\phi^2\phi^2 + \frac{\lambda_\phi}{4!}\phi^4 \\ &\quad + \frac{1}{64\pi^2}\left(\frac{\lambda_\phi^2\phi^4}{4}(\ln\frac{M_\phi^2}{\mu^2} - I(M_\phi^2/\mu^2)) + g^4\phi^4(\ln\frac{M_\chi^2}{\mu^2} - I(M_\chi^2/\mu^2))\right)\end{aligned} \tag{C.11}$$

where $\lambda_\phi \equiv \lambda_\phi(\mu)$ and $g^2 \equiv g^2(\mu)$.

References

[1] L. F. Abbott, Edward Farhi, and Mark B. Wise. Particle Production in the New Inflationary Cosmology. *Phys. Lett.*, B117:29, 1982.

[2] Lotty Ackerman, Matthew R. Buckley, Sean M. Carroll, and Marc Kamionkowski. Dark Matter and Dark Radiation. *Phys. Rev.*, D79:023519, 2009.

[3] A. J. Albrecht, P. J. Steinhardt, M. S. Turner, and F. Wilczek. Reheating an inflationary universe. *Phys. Rev. Lett.*, 48:1437, 1982.

[4] B. Allen. The Stochastic Gravity Wave Background in Inflationary Universe Models. *Phys. Rev. D*, 37:2078, 1988.

[5] Luca Amendola. Coupled quintessence. *Phys. Rev.*, D62:043511, 2000.

[6] Claude Amsler et al. Review of particle physics. *Phys. Lett.*, B667:1–1340, 2008.

[7] Thomas Appelquist and J. Carazzone. Infrared Singularities and Massive Fields. *Phys. Rev.*, D11:2856, 1975.

[8] Alexandre Arbey and Farvah Mahmoudi. One-loop quantum corrections to cosmological scalar field potentials. *Phys. Rev.*, D75:063513, 2007.

[9] C. Armendariz-Picon, Viatcheslav F. Mukhanov, and Paul J. Steinhardt. A dynamical solution to the problem of a small cosmological constant and late-time cosmic acceleration. *Phys. Rev. Lett.*, 85:4438–4441, 2000.

[10] C. Armendariz-Picon, Viatcheslav F. Mukhanov, and Paul J. Steinhardt. Essentials of k-essence. *Phys. Rev.*, D63:103510, 2001.

[11] N. A. Bahcall and X. Fan. The Most Massive Distant Clusters: Determining Omega and delta 8. ApJ, 504:1, September 1998.

[12] M. Baldi. Time-dependent couplings in the dark sector: from background evolution to non-linear structure formation. MNRAS, 411:1077–1103, February 2011.

[13] R. D. Ball and R. S. Thorne. Renormalizability of effective scalar field theory. *Ann. Phys.*, 236:117–204, 1994.

[14] Richard D. Ball and Robert S. Thorne. The Decoupling theorem in effective scalar field theory. *Ann. Phys.*, 241:368–393, 1995.

[15] Masako Bando, Taichiro Kugo, Nobuhiro Maekawa, and Hiroaki Nakano. Improving the effective potential. *Phys. Lett.*, B301:83–89, 1993.

[16] Masako Bando, Taichiro Kugo, Nobuhiro Maekawa, and Hiroaki Nakano. Improving the effective potential: Multimass scale case. *Prog. Theor. Phys.*, 90:405–418, 1993.

[17] J. D. Barrow and T. Clifton. Cosmologies with energy exchange. *Physical Review D*, 73(10):103520, May 2006.

[18] Sergei Bashinsky and Edmund Bertschinger. Position-Space Description of the Cosmic Microwave Background and Its Temperature Correlation Function. *Phys. Rev. Lett.*, 87:081301, 2001.

[19] Mar Bastero-Gil and Arjun Berera. Warm inflation model building. *Int. J. Mod. Phys.*, A24:2207–2240, 2009.

[20] Mar Bastero-Gil, Arjun Berera, Brendan M. Jackson, and Andy Taylor. Hybrid Quintessential Inflation. *Phys. Lett.*, B678:157–163, 2009.

[21] R. Bean and O. Doré. Probing dark energy perturbations: The dark energy equation of state and speed of sound as measured by WMAP. Phys. Rev. D, 69(8):083503, April 2004.

[22] Jonathan Benjamin et al. Cosmological Constraints From the 100 Square Degree Weak Lensing Survey. *Mon. Not. Roy. Astron. Soc.*, 381:702–712, 2007.

[23] Arjun Berera. Warm Inflation. *Phys. Rev. Lett.*, 75:3218–3221, 1995.

[24] R. D. Blandford, Mustafa A. Amin, E. A. Baltz, K. Mandel, and P. J. Marshall. Cosmokinetics. *Observing Dark Energy (NOAO/Tucson proceedings)*, 2004.

[25] A. Bonanno, J. Polonyi, and D. Zappala. Renormalization group analysis of the heavy field decoupling in a scalar theory. *Como*, 1996.

[26] Philippe Brax and Jerome Martin. The robustness of quintessence. *Phys. Rev.*, D61:103502, 2000.

[27] Anthony W. Brookfield, C. van de Bruck, D. F. Mota, and D. Tocchini-Valentini. Cosmology of mass-varying neutrinos driven by quintessence: Theory and observations. *Phys. Rev.*, D73:083515, 2006.

[28] J. S. Bullock. Notes on the Missing Satellites Problem. *ArXiv e-prints*, September 2010.

[29] C. P. Burgess. Introduction to effective field theory. *Ann. Rev. Nucl. Part. Sci.*, 57:329–362, 2007.

[30] R. R. Caldwell, Rahul Dave, and Paul J. Steinhardt. Cosmological Imprint of an Energy Component with General Equation-of-State. *Phys. Rev. Lett.*, 80:1582–1585, 1998.

[31] A. H. Campos, H. C. Reis, and R. Rosenfeld. Preheating in quintessential inflation. *Phys. Lett.*, B575:151–156, 2003.

[32] S. Capozziello, V. F. Cardone, S. Carloni, and A. Troisi. Curvature quintessence matched with observational data. *Int. J. Mod. Phys.*, D12:1969–1982, 2003.

[33] S. M. Carroll, A. de Felice, V. Duvvuri, D. A. Easson, M. Trodden, and M. S. Turner. Cosmology of generalized modified gravity models. Phys. Rev. D, 71(6):063513, March 2005.

[34] J. A. Casas, V. Di Clemente, and M. Quiros. The effective potential in the presence of several mass scales. *Nucl. Phys.*, B553:511–530, 1999.

[35] M. Chevallier and D. Polarski. Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D*, 10:213–223, 2001.

[36] Takeshi Chiba. 1/R gravity and scalar-tensor gravity. *Phys. Lett.*, B575:1–3, 2003.

[37] Takeshi Chiba, Takahiro Okabe, and Masahide Yamaguchi. Kinetically driven quintessence. *Phys. Rev.*, D62:023511, 2000.

[38] Luis P. Chimento. Form invariance of differential equations in general relativity. *Journal of Mathematical Physics*, 38:2565, 1997.

[39] Sirichai Chongchitnan. Cosmological Perturbations in Models of Coupled Dark Energy. *Phys. Rev.*, D79:043522, 2009.

[40] J. M. Chung and B. K. Chung. Renormalization group improvement of the effective potential in massive phi**4 theory. *Phys. Rev.*, D60:105001, 1999.

[41] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. ApJ, 648:L109–L113, September 2006.

[42] S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *Mon. Not. Roy. Astron. Soc.*, 362:505–534, September 2005.

[43] Sidney R. Coleman and Erick J. Weinberg. Radiative Corrections as the Origin of Spontaneous Symmetry Breaking. *Phys. Rev.*, D7:1888–1910, 1973.

[44] John C. Collins. Scaling behavior of phi-to-the-4 theory and dimensional regularization. *Phys. Rev.*, D10:1213–1218, 1974.

[45] Edmund J. Copeland, Andrew R. Liddle, and James E. Lidsey. Steep inflation: Ending braneworld inflation by gravitational particle production. *Phys. Rev.*, D64:023509, 2001.

[46] Edmund J. Copeland, M. Sami, and Shinji Tsujikawa. Dynamics of dark energy. *International Journal of Modern Physics D*, 15:1753, 2006.

[47] P. S Corasaniti. Slow-Roll Suppression of Adiabatic Instabilities in Coupled Scalar Field-Dark Matter Models. *Physical Review D*, 78:083538, 2008.

[48] A. Dabelstein. The One loop renormalization of the MSSM Higgs sector and its application to the neutral scalar Higgs masses. *Z. Phys.*, C67:495–512, 1995.

[49] Antonio De Felice, Salah Nasri, and Mark Trodden. Quintessential baryogenesis. *Phys. Rev.*, D67:043509, 2003.

[50] Konstantinos Dimopoulos. The curvaton hypothesis and the eta-problem of quintessential inflation, with and without branes. *Phys. Rev.*, D68:123506, 2003.

[51] Antonio Dobado, Maria J. Herrero, W. Hollik, and Siannah Penaranda. Quantum effects to the Higgs boson self-couplings in the SM and in the MSSM, 2002.

[52] A. D. Dolgov and A. D. Linde. Baryon asymmetry in inflationary universe. *Physics Letters B*, 116:329, 1982.

[53] Michael Doran and Joerg Jaeckel. Loop corrections to scalar quintessence potentials, 2002.

[54] J. Dunkley, E. Komatsu, M. R. Nolta, D. N. Spergel, D. Larson, G. Hinshaw, L. Page, C. L. Bennett, B. Gold, N. Jarosik, J. L. Weiland, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data. ApJS, 180:306–329, February 2009.

[55] G. R. Dvali, G. Gabadadze, and M. Porrati. Metastable gravitons and infinite volume extra dimensions. *Phys. Lett.*, B484:112–118, 2000.

[56] G. R. Dvali, Q. Shafi, and Robert K. Schaefer. Large scale structure and supersymmetric inflation without fine tuning. *Phys. Rev. Lett.*, 73:1886–1889, 1994.

[57] M. B. Einhorn and D. R. T. Jones. The effective potential, the renormalisation group and vacuum stability. *Journal of High Energy Physics*, 4:51, April 2007.

[58] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function

of SDSS Luminous Red Galaxies. *Astrophysical Journal*, 633:560–574, November 2005.

[59] Glennys R. Farrar and P. James E. Peebles. Interacting Dark Matter and Dark Energy. *Astrophys. J.*, 604:1–11, 2004.

[60] Bo Feng and Ming-zhe Li. Curvaton reheating in non-oscillatory inflationary models. *Phys. Lett.*, B564:169–174, 2003.

[61] C. Ford, I. Jack, and D. R. T. Jones. The Standard Model Effective Potential at Two Loops. *Nucl. Phys.*, B387:373–390, 1992.

[62] C. Ford and D. R. T. Jones. The Effective potential and the differential equations method for Feynman integrals. *Phys. Lett.*, B274:409–414, 1992.

[63] C. Ford, D. R. T. Jones, P. W. Stephenson, and M. B. Einhorn. The Effective potential and the renormalization group. *Nucl. Phys.*, B395:17–34, 1993.

[64] L. H. Ford. Gravitational particle creation and inflation. *Phys. Rev. D*, 35:2955, 1987.

[65] Mathias Garny. Quantum corrections in quintessence models. *Phys. Rev.*, D74:043009, 2006.

[66] Howard Georgi and H. David Politzer. Freedom at Moderate Energies: Masses in Color Dynamics. *Phys. Rev.*, D14:1829, 1976.

[67] M. Giovannini. Relic gravity waves from brane world inflation. *Phys. Rev. D*, 58, 1998.

[68] Luciano Girardello and Alberto Zaffaroni. Exact renormalization group equation and decoupling in quantum field theory. *Nucl. Phys.*, B424:219–238, 1994.

[69] Tame Gonzalez, G. Leon, and I. Quiros. Dynamics of Quintessence Models of Dark Energy with Exponential Coupling to the Dark Matter. *Class. Quant. Grav.*, 23:3165–3179, 2006.

[70] S. Gorbunov, Dimitri and A. Rubakov, Valery. *Introduction to the Theory of the Early Universe: Cosmological Perturbations and Inflationary Theory.* World Scientific Publishing, 2010.

[71] Eric Greenwood, Evan Halstead, Robert Poltis, and Dejan Stojkovic. Dark energy, the electroweak vacua and collider phenomenology. *Phys. Rev.*, D79:103003, 2009.

[72] J.-H. He, B. Wang, and E. Abdalla. Stability of the curvature perturbation in dark sectors' mutual interacting models. *Physics Letters B*, 671:139–145, January 2009.

[73] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results. *ArXiv e-prints*, March 2008.

[74] Henk Hoekstra, Howard K. C. Yee, and Michael D. Gladders. Properties of galaxy dark matter halos from weak lensing. *Astrophys. J.*, 606:67–77, 2004.

[75] Wayne Hu. Acceleration from Modified Gravity: Lessons from Worked Examples. *Nucl. Phys. Proc. Suppl.*, 194:230–238, 2009.

[76] Greg Huey and James E. Lidsey. Inflation, braneworlds and quintessence. *Phys. Lett.*, B514:217–225, 2001.

[77] Dragan Huterer and Michael S. Turner. Probing the dark energy: Methods and strategies. *Phys. Rev.*, D64:123527, 2001.

[78] Brendan M. Jackson, Andy Taylor, and Arjun Berera. On the large-scale instability in interacting dark energy and dark matter fluids. *Phys. Rev.*, D79:043526, 2009.

[79] Vinod B. Johri and P. K. Rath. Parametrization of dark energy equation of state. *Int. J. Mod. Phys.*, D16:1581–1591, 2007.

[80] Michael Joyce. On the expansion rate of the universe at the electroweak scale. *Phys. Rev.*, D55:1875–1878, 1997.

[81] Manoj Kaplinghat and Arvind Rajaraman. Stable Models of superacceleration. *Phys. Rev.*, D75:103504, 2007.

[82] Boris M. Kastening. Renormalization group improvement of the effective potential in massive O(N) symmetric phi**4 theory. *UCLA-92-TEP-26*, 1992.

[83] Boris M. Kastening. Renormalization group improvement of the effective potential in massive phi**4 theory. *Phys. Lett.*, B283:287–292, 1992.

[84] Yoichi Kazama and York-Peng Yao. Decoupling, Effective Lagrangian, and Gauge Hierarchy in Spontaneously Broken Nonabelian Gauge Theories. *Phys. Rev.*, D25:1605, 1982.

[85] T. W. B. Kibble. Topology of Cosmic Domains And Strings. *J. Phys. A*, 9:1387, 1976.

[86] Lev Kofman, Andrei D. Linde, and Alexei A. Starobinsky. Reheating after inflation. *Phys. Rev. Lett.*, 73:3195–3198, 1994.

[87] Lev Kofman, Andrei D. Linde, and Alexei A. Starobinsky. Towards the theory of reheating after inflation. *Phys. Rev.*, D56:3258–3295, 1997.

[88] T. Koivisto. Growth of perturbations in dark matter coupled with quintessence. Phys. Rev. D, 72(4):043516, August 2005.

[89] Edward Kolb and Michael Turner. *The Early Universe*. Addison-Wesley Publishing Company, 1990.

[90] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18, February 2011.

[91] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18, February 2011.

[92] Martin Kunz and Domenico Sapone. Dark energy versus modified gravity. *Phys. Rev. Lett.*, 98:121301, 2007.

[93] B. Li and J. D. Barrow. On the effects of coupled scalar fields on structure formation. MNRAS, 413:262–270, May 2011.

[94] Baojiu Li and John D. Barrow. N-Body Simulations for Coupled Scalar Field Cosmology. *Phys. Rev.*, D83:024007, 2011.

[95] Andrew R. Liddle and L. Arturo Urena-Lopez. Curvaton reheating: An application to braneworld inflation. *Phys. Rev.*, D68:043517, 2003.

[96] A. D. Linde. Chaotic inflation. *Phys. Lett. B*, 129:177, 1983.

[97] Andrei D. Linde. A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems. *Phys. Lett.*, B108:389–393, 1982.

[98] Andrei D. Linde. Hybrid inflation. *Phys. Rev.*, D49:748–754, 1994.

[99] Eric V. Linder. Exploring the expansion history of the universe. *Phys. Rev. Lett.*, 90:091301, 2003.

[100] Eric V. Linder. Cosmic growth history and expansion history. *Phys. Rev.*, D72:043529, 2005.

[101] Eric V. Linder and Robert N. Cahn. Parameterized Beyond-Einstein Growth. *Astropart. Phys.*, 28:481–488, 2007.

[102] David H. Lyth and Antonio Riotto. Particle physics models of inflation and the cosmological density perturbation. *Phys. Rept.*, 314:1–146, 1999.

[103] C.-P. Ma and E. Bertschinger. Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. ApJ, 455:7, December 1995.

[104] Roy Maartens, David Wands, Bruce A. Bassett, and Imogen Heard. Chaotic inflation on the brane. *Phys. Rev.*, D62:041301, 2000.

[105] Marie E. Machacek and Michael T. Vaughn. Two Loop Renormalization Group Equations in a General Quantum Field Theory. 1. Wave Function Renormalization. *Nucl. Phys.*, B222:83, 1983.

[106] Marie E. Machacek and Michael T. Vaughn. Two Loop Renormalization Group Equations in a General Quantum Field Theory. 3. Scalar Quartic Couplings. *Nucl. Phys.*, B249:70, 1985.

[107] Aneesh V. Manohar. Effective field theories. 1996.

[108] J. C. Mather, E. S. Cheng, D. A. Cottingham, R. E. Eplee, Jr., D. J. Fixsen, T. Hewagama, R. B. Isaacman, K. A. Jensen, S. S. Meyer, P. D. Noerdlinger, S. M. Read, L. P. Rosen, R. A. Shafer, E. L. Wright, C. L. Bennett, N. W. Boggess, M. G. Hauser, T. Kelsall, S. H. Moseley, Jr., R. F. Sil, G. F. Smoot, R. Weiss, and D. T. Wilkinson. Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *Astrophysical Journal*, 420:439–444, January 1994.

[109] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. ApJ, 270:365–370, July 1983.

[110] Takahiro Miura, Eiichi Takasugi, and Masaki Yoshimura. Quantum effects for the neutrino mixing matrix in the democratic-type model. *Prog. Theor. Phys.*, 104:1173–1187, 2000.

[111] Moore, B. and Ghigna, S. and Governato, F. and Lake, G. and Quinn, J. Thomas R. and Stadel, J. and Tozzi, P. Dark Matter Substructure in Galactic Halos. *Astrophysical Journal Letters*, 1999.

[112] Hiroaki Nakano and Yuhsuke Yoshida. Improving the effective potential, multi-mass problem and modified mass dependent scheme. *Phys. Rev.*, D49:5393–5407, 1994.

[113] John Peacock. *Cosmological Physics*. 'Cambridge University Press, 2005.

[114] P. J. E. Peebles and Bharat Ratra. The cosmological constant and dark energy. *Rev. Mod. Phys.*, 75:559–606, 2003.

[115] P. J. E. Peebles and A. Vilenkin. Quintessential inflation. *Phys. Rev.*, D59:063505, 1999.

[116] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophysical Journal*, 142:419–421, July 1965.

[117] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer,

B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *Astrophysical Journal*, 517:565–586, June 1999.

[118] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Westview Press, 1995.

[119] Joseph Polchinski. Renormalization and Effective Lagrangians. *Nucl. Phys.*, B231:269–295, 1984.

[120] Joseph Polchinski. Effective field theory and the Fermi surface, 1992.

[121] M. Quartin, M. O. Calvão, S. E. Jorás, R. R. R. Reis, and I. Waga. Dark interactions and cosmological fine-tuning. *Journal of Cosmology and Astro-Particle Physics*, 5:7, May 2008.

[122] P Ramond. *Field Theory: a modern primer*. 'Addison-Wesley Publishing Co., 1990.

[123] Lisa Randall and Raman Sundrum. A large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[124] A. G. Riess, L.-G. Strolger, J. Tonry, S. Casertano, H. C. Ferguson, B. Mobasher, P. Challis, A. V. Filippenko, S. Jha, W. Li, R. Chornock, R. P. Kirshner, B. Leibundgut, M. Dickinson, M. Livio, M. Giavalisco, C. C. Steidel, T. Benítez, and Z. Tsvetanov. Type Ia Supernova Discoveries at $z > 1$ from the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy Evolution. *Astrophysical Journal*, 607:665–687, June 2004.

[125] Lewis. H Ryder. *Quantum field theory*. Cambridge University Press, 2006.

[126] Varun Sahni. The energy density of relic gravity waves from inflation. *Phys. Rev.*, D42:453–463, 1990.

[127] M. Sami, N. Dadhich, and Tetsuya Shiromizu. Steep inflation followed by Born-Infeld reheating. *Phys. Lett.*, B568:118–126, 2003.

[128] M. Sami and V. Sahni. Quintessential inflation on the brane and the relic gravity wave background. *Phys. Rev.*, D70:083513, 2004.

[129] T. Schimert and C. C. Chiang. Heavy Mass Power Counting Theorem and a Proof of Decoupling in the Momentum Subtraction Scheme. *Phys. Rev.*, D29:241, 1984.

[130] Vedat Nefer Senoguz and Qaisar Shafi. Chaotic inflation, radiative corrections and precision cosmology. *Phys. Lett.*, B668:6–10, 2008.

[131] Q. Shafi and Vedat Nefer Senoguz. Coleman-Weinberg Potential In Good Agreement With WMAP. *Phys. Rev.*, D73:127301, 2006.

[132] Fergus Simpson. Redshift Sensitivity of the Kaiser Effect. *Phys. Rev.*, D81:043513, 2010.

[133] Fergus Simpson. Scattering of Dark Matter and Dark Energy. *Phys. Rev.*, D82:083505, 2010.

[134] Parampreet Singh, M. Sami, and Naresh Dadhich. Cosmological dynamics of phantom field. *Phys. Rev.*, D68:023522, 2003.

[135] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. *Astrophys. J.*, 396:L1–L5, September 1992.

[136] B. Spokoiny. Deflationary universe scenario. *Phys. Lett. B*, 315:40, 1993.

[137] A. A. Starobinsky. Spectrum of relic gravitational radiation and the early state of the universe. *JETP Lett.*, 30:682, 1979.

[138] Dejan Stojkovic, Glenn D. Starkman, and Reijiro Matsuo. Dark energy, the colored anti-de Sitter vacuum, and LHC phenomenology. *Phys. Rev.*, D77:063006, 2008.

[139] K. Symmanzik. *Comm. Math. Phys*, 34:7, 1973.

[140] M. Tegmark et al. Cosmological parameters from SDSS and WMAP. Phys. Rev. D, 69(10):103501, May 2004.

[141] J. A. Tyson, R. A. Wenk, and F. Valdes. Detection of systematic gravitational lens galaxy image alignments - Mapping dark matter in galaxy clusters. *Astrophys. J.*, 349:L1–L4, 1990.

[142] M. Sami V. Sahni and T. Souradeep. Relic gravity waves from brane world inflation. *Phys. Rev. D*, 65, 2002.

[143] Jussi Valiviita, Roy Maartens, and Elisabetta Majerotto. Observational constraints on an interacting dark energy model. *Mon. Not. Roy. Astron. Soc.*, 402:2355–2368, 2010.

[144] Jussi Valiviita, Elisabetta Majerotto, and Roy Maartens. Instability in interacting dark energy and dark matter fluids. *JCAP*, 0807:020, 2008.

[145] A Vilenkin and E. P. Shellard. *Cosmic Strings and Other Topological Defects.* Cambridge University Press, 1994.

[146] Li-Min Wang and Paul J. Steinhardt. Cluster Abundance Constraints on Quintessence Models. *Astrophys. J.*, 508:483–490, 1998.

[147] Peng Wang and Xin-He Meng. Can vacuum decay in our universe? *Class. Quant. Grav.*, 22:283–294, 2005.

[148] Steven Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* John Wiley Sons Inc, 1972.

[149] Steven Weinberg. Perturbative Calculations of Symmetry Breaking. *Phys. Rev.*, D7:2887–2910, 1973.

[150] Steven Weinberg. The cosmological constant problem. *Rev. Mod. Phys.*, 61:1–23, 1989.

[151] Jochen Weller and Andreas Albrecht. Opportunities for future supernova studies of cosmic acceleration. *Phys. Rev. Lett.*, 86:1939–1942, 2001.

[152] C. Wetterich. Cosmology and the Fate of Dilatation Symmetry. *Nucl. Phys.*, B302:668, 1988.

[153] S. D. M. White, G. Efstathiou, and C. S. Frenk. The amplitude of mass fluctuations in the universe. *Mon. Not. Roy. Astron. Soc.*, 262:1023–1028, June 1993.

[154] S. D. M. White, J. F. Navarro, A. E. Evrard, and C. S. Frenk. The baryon content of galaxy clusters: a challenge to cosmological orthodoxy. Nature, 366:429–433, December 1993.

[155] K. G. Wilson and John B. Kogut. The Renormalization group and the epsilon expansion. *Phys. Rept.*, 12:75–200, 1974.

[156] Kenneth G. Wilson. Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture. *Phys. Rev.*, B4:3174–3183, 1971.

[157] Winfried Zimdahl and Diego Pavon. Interacting quintessence. *Phys. Lett.*, B521:133–138, 2001.